Exploration 1.3: Can dogs understand human cues?

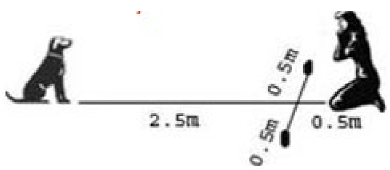
Alternative Measure of Strength of Evidence

LEARNING GOALS

- Find a standardized statistic from the observed proportion of “successes,” the hypothesized mean, and the SD of the null distribution as produced by the One Proportion applet.
- Interpret a standardized statistic
- State a conclusion about the alternative hypothesis (and null hypothesis) based on the magnitude of the standardized statistic
- Recognize that standardizing the statistic is an alternative measure of strength of evidence

Dogs have been domesticated for about 14,000 years. In that time, have they been able to develop an understanding of human gestures such as pointing or glancing? How about similar nonhuman cues? Researchers Udell, Giglio, and Wynne tested a small number of dogs in order to explore these questions.

In this exploration, we will first see whether dogs can understand human gestures as well as nonhuman gestures. To test this, the researchers positioned the dogs about 2.5 meters from the experimenter. Two cups were placed, one on each side of the experimenter.



The experimenter would perform some sort of gesture (pointing, bowing, looking) toward one of the cups or there would be some other nonhuman gesture (a mechanical arm pointing, a doll pointing, or a stuffed animal looking) toward one of the cups. The researchers would then see whether the dog would go to the cup that was indicated by the gesture. There were six dogs tested. We will look at one of the dogs in two of his sets of trials. This dog, a four-year-old mixed breed, was named Harley. Each trial involved one gesture and one pair of cups, with a total of 10 trials in a set.

We will start out by looking at one set of trials where the experimenter bowed toward one of the cups to see whether Harley would go to that cup.

STEP 1: State the research question.

1. Based on the description of the study, state the research question.

STEP 2: Design a study and collect data.

Harley was tested 10 times and 9 of those times he chose the correct cup.

2. What are the observational units in this study? (*Hint*: It’s not Harley, what is the sample size?)
3. Identify the variable in the study. What are possible outcomes of this variable? Is the variable quantitative or categorical?

The parameter of interest here is the probability that Harley chooses the correct cup in any one attempt.

4. State the null and alternative hypotheses to be tested. Express these both in words and symbols. (*Hint*: Think about the parameter and the research question of interest here.)

STEP 3: Explore the data.

With categorical data, we typically report the number of "successes" or the proportion of successes as the statistic.

5. What proportion of trials was Harley successful?

When we conduct analyses with binary variables, we often call one of the outcomes a "success" and the other a "failure" and then focus the analysis on the "success" outcome. It is arbitrary which outcome is defined to be a success, but you need to make sure you do so consistently throughout the analysis. For example, if you are looking at a survival rate after an operation, you will focus on survive(yes) as a success; however, if you are looking at a death rate after an operation, you will focus on survive(no) as a success.

STEP 4: Draw Inferences.

You will use the **One Proportion** applet to investigate how surprising Harley's observed statistic would be if he were just randomly selecting which cup to go to.

6. Before you use the applet, indicate what you will enter for the following values to match the head bowing study:
   a. Probability of success:
   b. Sample size:
   c. Number of repetitions:
7. Conduct this simulation analysis. Make sure the **Proportion of heads** button is selected in the applet and not **Number of heads**.
   a. Indicate how to calculate the approximate p-value (count the number of simulated statistics that equal ____ or ___________).
   b. Report the approximate p-value.
   c. Use the p-value to evaluate the strength of evidence provided by the sample data against the null hypothesis, in favor of the alternative that Harley really does understand bowing.

The p-value is the most common way to evaluate strength of evidence against the null hypothesis, but now we will explore a common alternative way to evaluate strength of evidence. The goal of any measure of strength of evidence is to use a number to help us determine whether the observed statistic falls in the tail of the null distribution (and is therefore surprising when the null hypothesis is true) or among the typical values we see when the null hypothesis is true.

8. Check the **Summary Statistics** box in the applet.
   a. Report the mean (average) value of the simulated statistics.
   b. Explain why it makes sense that this mean is close to 0.50.
   c. Report the standard deviation (SD) of the simulated statistics.
   d. Recall the observed value of the statistic. (What proportion of trials did Harley go to the correct cup?)

$$\hat{p} =$$

e. Calculate how many standard deviations the observed value of the statistic is from the hypothesized mean of the null distribution, 0.50. In other words, subtract the 0.50 from the observed value and then divide by the standard deviation. That is, calculate:

$$(observed\ statistic\ (\hat{p}) - 0.50)/SD\ of\ the\ null\ distribution$$

Your calculation in #8e is called "standardizing the statistic." It is telling us how far above the mean the observed statistic is in terms of the "how many standard deviations." This is how we will measure distance across different distributions, by standardizing and putting the observations on a common scale.

Definition

To standardize a statistic, compute the distance of the statistic from the (hypothesized) mean of the null distribution and divide by the standard deviation of the null distribution.

$$standardized\ statistic = \frac{statistic - mean\ of\ null\ distribution}{standard\ deviation\ of\ null\ distribution}$$

Once you calculate this value, you interpret it as "how many standard deviations the observed statistic falls from the hypothesized parameter value."

The next question is how to evaluate strength of evidence against the null hypothesis based on a standardized value. Here are some guidelines:

Guidelines for evaluating strength of evidence from standardized values of statistics

Standardizing gives us a quick, informal way to evaluate the strength of evidence against the null hypothesis. For standardized statistics:

between −1.5 and 1.5	**little or no** evidence against the null hypothesis
below −1.5 or above 1.5	**moderate** evidence against the null hypothesis
below −2 or above 2	**strong** evidence against the null hypothesis
below −3 or above 3	**very strong** evidence against the null hypothesis

The diagram in **Figure 1.3.4** illustrates the basis for using a standardized statistic to assess strength of evidence against the null hypothesis for a mound-shaped, symmetric distribution.

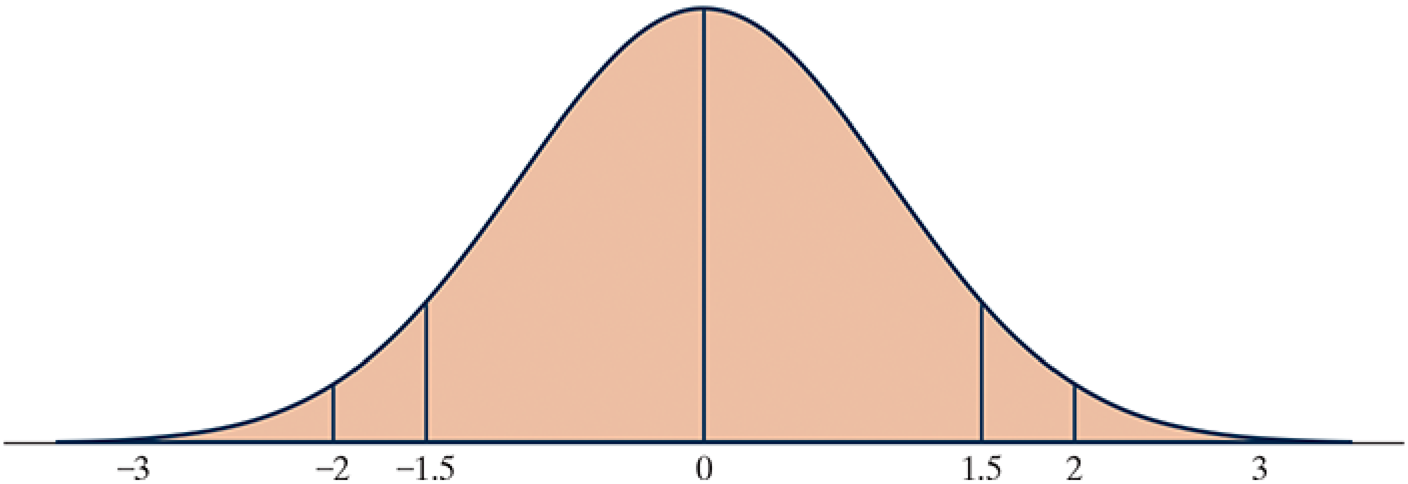


FIGURE 1.3.4 Positions of standardized statistics for a bell-shaped distribution.

The figure can be summarized by the following key idea.

> **Key Idea**
> Observations that fall more than 2 or 3 standard deviations from the mean can be considered in the tail of the distribution.

STEP 5: Formulate conclusions.

9. Let's examine the strength of evidence against the null.
    a. Based on the value of the standardized statistic, *z*, in #8e and the guidelines shown above, how much evidence does Harley's results provide against the null hypothesis?
    b. How closely does your evaluation of strength of evidence based on the standardized statistic compare to the strength of evidence based on the p-value in #7c?

Keep in mind that we haven't examined "additional evidence" against the null hypothesis, but rather an alternative way to measure the strength of evidence. The conclusions should always be consistent with each other.

Now, let's step back a bit further and think about the scope of inference. Based on the findings in this study, do you think this means that dogs understand bowing? Furthermore, does this mean that all dogs understand human cues?

STEP 6: Look back and ahead.

10. Based on the limitations of this study, suggest a new research question that you would investigate next.

Exploring Further

11. In #5 you recorded the proportion of trials where Harley went to the correct cup. Imagine that the proportion was actually larger (that Harley went to the correct cup 10 out of the 10 trials).

a. How would this have affected the p-value:

Larger | Same | Smaller

b. How would this have affected the absolute value of the standardized statistic:

Larger | Same | Smaller

c. How would this have affected the strength of evidence against the null hypothesis:

Stronger | Same | Weaker

12. Suppose that Harley went to the correct cup less than half of the trials, so the study result was in the opposite direction of the research conjecture and the alternative hypothesis.
    a. What can you say about the standardized value of the statistic in this case? Explain. (*Hint*: You cannot give a value for the standardized statistic, but you can say something specific about its value.)
    b. What can you say about the strength of evidence against the null hypothesis and in favor of the alternative hypothesis in this case?