Exploration 1.2: Can dogs smell COVID?

Measuring the Strength of Evidence

LEARNING GOALS

- Use appropriate symbols for parameter and statistic.
- State the null and alternative hypotheses in words and in terms of the symbol π, the long-run proportion.
- Explain how to conduct a simulation using a null hypotheses probability that is not 50-50.
- Use the **One Proportion** applet to obtain the p-value after carrying out an appropriate simulation.
- Anticipate the location of the center of the null distribution and how it changes based on whether you are using proportion or count as the statistic.
- Interpret the p-value.
- Explain why a smaller p-value provides stranger evidence against the null hypothesis.
- State a conclusion about the alternative hypothesis and null hypothesis based on the p-value.

To test whether or not a chance model of equally choosing between two options is plausible based on some observed data, we can flip a coin. However, not all situations call for a coin-flipping model. The coin-flipping model would work well to simulate guessing on a true-false test with a 50% chance of getting a question correct, but what about a multiple-choice test where you are guessing between four possible answers with a 25% chance of getting a question correct? We will apply the 6-step statistical investigation method in the more general setting of an arbitrary probability of "success."

We will also learn helpful terminology that is commonly used in statistical investigations to describe the process of drawing conclusions from data. We will work toward formalizing the procedure of a *test of significance* and give some guidelines to help determine when we have strong evidence that our chance model is not correct. We will also introduce some symbols, for convenience, but the big picture of assessing evidence against a claim is the same.

We will look at a study that tests dog's sense of smell. Dogs have a keen sense of smell. They are used for search and rescue, explosive detection, sniffing out illegal drugs in luggage at airports, and locating game while hunting. Can they also tell whether someone has COVID-19 by sniffing a specimen of sweat from a person? We will be looking at a study that used several dogs to test this question. We will focus on one dog, Maika, a 3-year-old female Belgian Malinois whose specialty is search and rescue. Maika completed 57 trials where she would sniff four different sweat specimens, one of which was from a COVID positive person, and then sit in front of the specimen she determined to be the positive specimen.

STEP 1: State the research question.

1. What is the research question that the researchers hoped to answer?

STEP 2: Design a study and collect data.

2. Identify the observational units in this study.
3. Identify the variable. Is the variable quantitative of categorical?

Definition
A binary variable is a categorical variable with only two outcomes. Often we convert categorical variables with more than two outcomes (e.g., blood type: A, B, AB, O) into binary variables (e.g., type A or not type A). In this case, we also define one of the outcomes to be a "success" and one to be a "failure."

4. Write the variable you identify in #3 as a binary variable.
5. Describe the parameter of interest in this study (in words). (*Hint*: The parameter of interest is the long-run proportion of …?)
6. One possibility here is that Maika can't smell COVID and is equally likely to choose any of the four scent specimens as the COVID positive specimen, essentially selecting one of the four specimens at random. In this case, what is the long-run proportion (i.e., probability) that Maika selects the COVID positive specimen in any particular attempt?
7. Another possibility is that Maika can smell COVID and is more likely to select the COVID positive specimen than if she was randomly guessing. In this case, what can you say about the long-run proportion of times Maika selects the COVID positive specimen? (*Hint*: You are not to specify a particular value at this time, instead indicate a direction from a particular value.)

Definitions
The ***null hypothesis*** typically represents the "by-random-chance-alone" explanation. The chance model (or "null model") is chosen to reflect this hypothesis.
The ***alternative hypothesis*** typically represents the "there is an effect" explanation that contradicts the null hypothesis. Researchers typically hope this hypothesis will be supported by the data they collect.

8. Your answers to #6 and #7 should be the null and alternative hypotheses for this study. Which is which?

STEP 3: Explore the data.

The researchers found that in 47 of the 57 trials Maika chose the COVID positive specimen.

9. Calculate the value of the relevant statistic.

Use of Symbols

We can use mathematical symbols to represent quantities and simplify our writing. Throughout the book we will emphasize written explanations but will also show you mathematical symbols which you are free to use as a short-hand once you are comfortable with the material. The distinction between parameter and statistic is so important that we always use different symbols to refer to them.

When dealing with a parameter that is a long-run proportion, such as the probability that Maika would choose the COVID positive specimen, we use the Greek letter π (pronounced "pie"). But when working with a statistic that is the proportion of "successes" in a sample, such as the proportion of trials Maika *did* choose the OCVID positive sample, we use the symbol (pronounced "p-hat"). Finally, we use the symbol n to represent the sample size.

10. What is the value of $\hat{p}$ in this study?
11. What is the value of n in this study?

12. Hypotheses are always conjectures about the unknown parameter. You can also use H_0 and H_a as short-hand notation for the null and alternative hypotheses, respectively. A colon, ":", is used to represent the word "is." Restate the null and alternative hypotheses using π to represent the unknown probability that Maika will choose the positive specimen

    H_0:

    H_a:

STEP 4: Draw inferences.

13. Is the sample proportion of correct identifications in this study larger than the probability specified in the null hypothesis?
14. Is it possible that this proportion could turn out to be this large even if the null hypothesis was true? (i.e., even if Maika couldn't smell COVID and was essentially selecting at random from the four specimens)?

We will use simulation to investigate how surprising the observed sample result (47 of 57 correct COVID identifications) would be if in fact Maika could not smell COVID and so for each trial had a 0.25 probability of selecting the COVID specimen. (Note also that our null model assumes the same probability for each trial.)

Think About It

Can we use a single coin toss to represent the chance model specified by the null hypothesis? If not, can you suggest a different random device that we could use? What needs to be different about our simulation?

15. Explain why we cannot use a simple coin toss to simulate Maika's choices, as we did with a 50-50 chance of success.
16. We could do the simulation using a set of four playing cards: one black and three red. Explain how the simulation would work in this case.
17. Another option would be to use a spinner like the one shown here,  like you would use when playing a child's board game. Explain how the simulation would work if you were using a spinner. In particular:
    a. What does each region represent?
    b. How many spins of the spinner will you need to do in order to simulate one repetition of the experiment when there is equal preference between the four specimens (null hypothesis is true)?
18. We will now use the **One Proportion** applet to conduct this simulation analysis. Notice that the applet will show us what it would be like if we were simulating with spinners.
    a. First enter the **probability of heads/probability of success** value specified in the null hypothesis.
    b. Enter the appropriate **sample size** (number of Maika's trials in this study).

c. Keep 1 for the number of samples, and press **Draw Samples**. Report the number of "successes" in this simulated sample.
d. Now, select the radio button for "Proportion of successes." What value on the Proportion of successes graph is this simulated sample proportion of success close to? Use your answer to "c" to verify how this simulated value is calculated.
e. Leaving the "Proportion of successes" radio button selected, click on **Draw Samples** four more times. Do you get the same results each time?
f. Now enter 995 for the number of samples and click on **Draw Samples**, bringing the number of simulated samples to 1,000. Comment on the center, variability, and shape of the resulting distribution of sample proportions.

This distribution of simulated sample proportions is called the ***null distribution***, because it is created assuming the null hypothesis to be true.

19. Recall that the observed value of the sample proportion of correctly identified COVID specimens in this study was = 47/57 ≈ 0.83. Looking at the null distribution you have simulated, is this a very unlikely result when the null hypothesis is true? In other words, is this value far in the tail of the null distribution?

In this case, the observed statistic is far out in the tail of the distribution and it is not hard to see that Maika's proportion of successful identifications is unlikely to happen by random chance. There will be studies when the observed statistic is not that far in the tail of the distribution, but also not near the middle of the distribution (e.g., what if Maika had been correct 35% of the time). To help make a judgement about strength of evidence in this case, we can count how many (and what proportion) of the simulated sample proportions are as extreme or more extreme than the observed value.

20. Use the applet to count how many (and what proportion) of the simulated sample proportions are as or more extreme than the observed value. Make sure that the ≥ inequality symbol is selected (to match the alternative hypothesis). Then enter **0.83** (the observed sample proportion of correct COVID positive identifications) in the box to the left of the **Count** button. Then click on the **Count** button. Record the number and proportion of simulated sample proportions that are as extreme or more extreme than the observed value.

Definition
The ***p-value*** is the probability of obtaining a value of the statistic at least as extreme as the observed statistic when the null hypothesis is true. We can estimate the *p*-value by finding the proportion of the simulated statistics in the null distribution that are *at least as extreme* (in the direction of the alternative hypothesis) as the value of the statistic actually observed in the research study.

How do we *evaluate* this *p*-value as a judgment about strength of evidence provided by the sample data against the null hypothesis? One answer is: The smaller the p-value, the stronger the evidence against the null hypothesis and in favor of the alternative hypothesis. But how small is small enough to regard as convincing? There is no definitive answer, but here are some guidelines:

Guidelines for evaluating the strength of evidence from p-values

0.10 < p-value	not much evidence against null hypothesis; null is plausible
0.05 < p-value ≤ 0.10	moderate evidence against the null hypothesis

0.01 < p-value ≤ 0.05	strong evidence against the null hypothesis
p-value ≤ 0.01	very strong evidence against the null hypothesis

The smaller the p-value, the stronger the evidence against the null hypothesis.

21. Is the approximate p-value from your simulation analysis (your answer to #20) small enough to provide convincing evidence against the null hypothesis that Maika was just guessing which of the four specimens was COVID positive? If so, how strong is this evidence? Explain.
22. When computing p-values, "more extreme" is always measured in the direction of the alternative hypothesis. Use this fact to explain why you counted ≥ 0.83 in #20.

STEP 5: Formulate conclusions.

23. Do you consider the observed sample result to be *statistically significant*? Recall that this means that the observed result is unlikely to have occurred by chance alone.
24. How broadly are you willing to generalize your conclusions? Would you be willing to generalize your conclusions to all dogs? Explain your reasoning.

STEP 6: Look back and ahead.

25. Suggest a new research question that you might investigate next, building on what you learned in this study.

Exploring Further

Instead of focusing on Maika's successful identifications, you could instead analyze the data based on her incorrect identifications. Because Maika correctly identified the COVID positive specimen in 47 of 57 trials, we know that she was incorrect in her identification of the COVID positive specimen in 10 of the 57 trials. Now let the parameter of interest be the probability that Maika will be incorrect (again denoted by π).

26. Conduct a simulation analysis to assess the strength of evidence provided by the sample data.
    a. The research conjecture is that Maika tends to identify the incorrect specimen (*more* or *less*) often than random chance. (Circle your answer.)
    b. State the null hypothesis in words and in terms of the (newly defined) parameter π.
    c. State the alternative hypothesis in words and in terms of the (new) parameter π.
    d. Calculate the observed value of the relevant statistic.
    e. Before you use the **<u>One Proportion</u>** applet to analyze these data, indicate what values you will input:

       Probability of success:

       Sample size:

       Number of samples:

    f. Use the applet to produce the null distribution of simulated sample proportions. Comment on the center, variability, and shape of this distribution. Be sure to comment on how this null distribution differs from the null distribution in #18(f).
    g. In order to approximate the p-value, you will count how many of the simulated proportions are _____ or _____ (larger/smaller) and then divide by _____.
    h. Estimate the p-value from your simulation results.

i. *Interpret* this p-value. (*Hint*: This is the probability of what, assuming what?)
j. *Evaluate* this p-value: How much evidence do the sample data provide against the null hypothesis?

27. Does your analysis based on the number of trials Maika was incorrect in her identification produce similar conclusions to your previous analysis based on the number of trials where Maika was correct? Explain.

You should have found that it does not matter whether you focus on the number/proportion of correct identifications or the number/proportion of incorrect identifications. In other words, it does not matter which category you define to be a "success" for the variable. Your findings should be very similar provided that you make the appropriate adjustments in your analysis:

- Using 0.75 instead of 0.25 as the null value of the parameter
- Changing the alternative hypothesis to "$\pi < 0.75$" rather than "$\pi > 0.25$"
- Calculating the p-value as the proportion of samples with as extreme as ≤ 0.17 rather than ≥ 0.83