Exploration 1.5: Eye Dominance?

Inference for a Single Proportion: Theory-based Test

LEARNING GOALS

- Predict the shape, mean, and standard deviation of the null distribution of sample proportions, as approximately normal (when the sample size is large enough), π, and $\sqrt{\pi(1-\pi)/n}$, respectively.
- Use the One Proportion applet to carry out the one-proportion *z*-test (theory-based; normal approximation-based), including estimating the p-value, and standardized statistic, *z*.
- Explain when simulation and theory will yield different answers.

Just like handedness where people prefer to use one hand over another, eye dominance, sometimes called eyedness, is the tendency to prefer to see using one eye over the other. Interestingly though the side of the dominant eye does not always match that of the dominant hand. Let's investigate whether people are equally likely to have left-eye or right-eye dominance by collecting some data from you and your classmates.

1. To figure out which of your eyes is the dominant eye, carry out the following "dominant eye test":
   - Extend both your arms in front of you and create a triangular opening using your thumbs and pointer fingers.
   - With both eyes open, center your triangular opening on a distant object such as a clock or a poster on the wall.
   - Close your left eye. If the object stays centered in your triangular opening, your right eye (as that is the one that's open) is your dominant eye. If the object is no longer in the triangular opening, your left eye is your dominant eye.
   - Double check this by closing your right eye. If the object stays centered in your triangular opening, your left eye (as that is the one that's open) is your dominant eye. If the object is no longer in the triangular opening, your right eye is your dominant eye.

   Record whether you have left-eye or right-eye dominance.

   Before we combine your data with the data from your classmates, let's think about what we want to test here. Conventional wisdom says that more often people are right-handed than left, so for now let's use our research hypothesis to be that more often people tend to have right-eye dominance than left-eye dominance.
2. In this study with your class:
   a. What are the observational units in this study?
   b. What is the variable that is recorded?
   c. Describe the parameter of interest in words. (Use the symbol π to represent this parameter.)
   d. If right-eye and left-eye dominance are equally prevalent, what would you expect the numerical value of the parameter to be? Is this the null hypothesis or the alternative hypothesis?
   e. If people are more likely to be right-eye dominant than left, what can you say about the numerical value of the parameter? Is this the null hypothesis or the alternative hypothesis?

3. Including yourself and your classmates, how many people participated in this study? How many have right-eye dominance? Calculate the sample proportion who are righteye dominant.
4. To have a larger sample size to analyze, combine your class results with the results from some of the authors' classes, in which 70 of 115 students had right-eye dominance. Now what are the sample size and the sample proportion that are right-eye dominant?
   Sample size:
   Sample proportion: (Use the symbol $\hat{p}$ or $\hat{p}_{observed}$ to represent this statistic.)

5. Use the **One Proportion** applet to test the hypotheses from #2d and #2e.
   a. Describe the shape of the null distribution of sample proportions. Does this shape look familiar? Where is the null distribution centered? Does this make sense? Check the **Summary Statistics** box and report the mean and standard deviation as reported by the applet.

   Shape: Familiar?
   Center? Why does this make sense?
   Mean: SD:

   b. Approximate the p-value and summarize the strength of evidence that the sample data provide regarding the research hypothesis.
   c. Determine the standardized statistic and confirm that the strength of evidence obtained using the standardized statistic is similar to that obtained using the p-value.

Theory-based Approach (One-Proportion z-test)

In #5a, you probably described the shape of the null distribution using words such as bell-shaped, symmetric, or maybe even approximately normal. You have seen many null distributions in this chapter that have had this same basic shape. You should have also noticed that the null distributions have all been centered at the hypothesized value of the long-run proportion used in the null hypotheses. Based on these earlier observations, you probably could have predicted that your null distribution was going to be somewhat bell-shaped and centered at 0.50. You probably would have a harder time predicting your null distribution's variability (standard deviation), but this too can be predicted in advance, as we will see shortly.

We can often use mathematical models known as normal distributions (bell-shaped curves) to approximate a null distribution of sample proportions. When such mathematical models are used to predict what the value of the standardized statistic and p-value, we call the approach a *theory-based approach*. The normal distribution provides a second way, in addition to simulation, to approximate a p-value.

6. Check the box next to **Normal Approximation** in the applet. Does the region shaded in blue seem to be a good description (model) of what we actually got in the simulation?

Validity Conditions for Theory-based Approach

The normal approximation to the null distribution is valid whenever the sample size is reasonably large. One convention is to consider the sample size to be large enough whenever there are at least 10 observations in each category (at least 10 successes and at least 10 failures).

7. According to this convention, is the sample size large enough in this study to use the normal approximation and theory-based inference? Justify your answer.

Validity Conditions
The normal approximation for the distribution of sample proportions can be thought of as a prediction of what would occur if a simulation-based analysis was carried out. Many times this prediction is valid, but not always. It is only valid when the condition (at least 10 successes and at least 10 failures) is met. This mathematical prediction of the shape of the distribution of sample proportions is often called the "central limit theorem."

Formulas

A mathematical approximation for the standard deviation of the distribution of sample proportions also exists. The standard deviation can be obtained in either of two ways:

One, find the standard deviation of the null distribution by simulating.
Two, predict the value of the standard deviation by substituting into this formula: $\sqrt{\pi(1-\pi)/n}$

8. Use the formula to predict the (theoretical) standard deviation of the distribution of sample proportions. Then compare this value to the SD of your simulated sample proportions, as recorded in #5a. Are they similar?

The predicted value of the standard deviation (using the formula) will be very close to the simulated standard deviation of the null distribution.

The mean of the null distribution is the hypothesized value of the long-run proportion (π).

Theory-based prediction of distribution of sample proportions
If the sample size, n, is large enough, the distribution of sample proportions will be bell-shaped (or normal), centered at the long-run proportion (π), with standard deviation of $\sqrt{\pi(1-\pi)/n}$.

Recall that the standardized score is calculated as

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

The symbol *z* is often used to represent this calculation, especially when the value will be compared to the standard normal distribution.

9. Use the predicted value of the standard deviation from #8 to calculate the standardized statistic by hand and confirm that your answer is very close to what you found in #5c when using simulation.

In the applet, see that the predicted value of the standardized statistic, *z*, is given immediately below the button for "Normal approximation" in parentheses and should match your answer to #9.

10. The theory-based (normal approximation) p-value is also displayed. Compare this p-value to the one you got from simulation (#5b). Are they similar?
11. Is there pretty good agreement between the standard deviation (#8), standardized statistic (#9) and p-value (#10) you calculated when using the theory-based (one-proportion *z*-test; normal approximation) to what you got in your simulation? Explain why you could have predicted this close alignment in advance of the theory-based calculations?

Exploring Further

Follow-up Analysis #1 There are several research papers (see, for example, Ehrenstein et al., 2005) on eye preference that say that the long-run proportion of right-eyed people is two-thirds.

12. Use the theory-based approach to test the claim that there is a two-thirds chance that a particular person will be right-eye dominant, using the sample data (your class combined with author's class) and a two-sided alternative. Report the null and alternative hypothesis, standardized statistic, and p-value. Summarize your conclusion and explain the reasoning process by which it follows from your analysis.

Follow-up Analysis #2 In a small class of 14 students, nine students turned out to be right-eye dominant.

13. Use simulation to generate a two-sided p-value evaluating the strength of evidence that the long-run proportion of right-eyed students is different than 50% based on this small class's data alone.
14. Why can't you use the normal approximation in this case?
15. Use the normal approximation anyway. Compare and comment on the p-values obtained from the two methods.