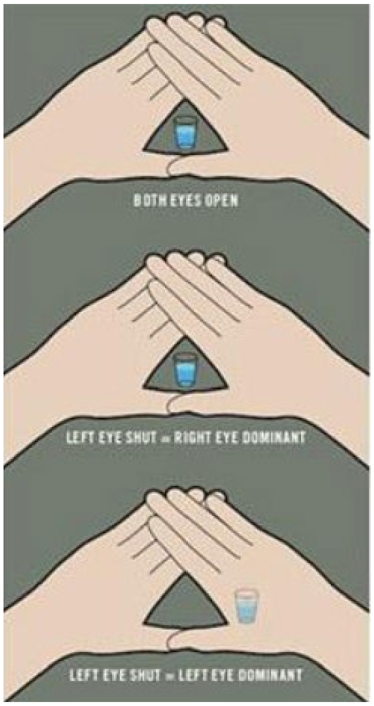
Exploration 3.4: Dominant Eye

Factors that affect the width of a confidence interval

LEARNING GOALS

- Recognize that all other things being equal, as the confidence level increases, the width of the confidence interval increases.
- Apply the idea that the confidence level of an interval corresponds to its coverage probability (the long-run proportion of confidence intervals containing the true parameter value across many, many random samples) in the interpretation of confidence intervals.
- Recognize that all other things being equal, as the sample size increases, the width of the resulting confidence interval decreases.
- Recognize that all other things being equal, as the standard deviation of a quantitative variable increases, the resulting confidence interval will be wider.
- Recognize that all other things being equal, as the sample proportion gets farther from 0.5, the standard error decreases and thus a resulting confidence interval will be narrower.
- Explain why a statistically significant result does not imply evidence of a large difference from what the null says—recall that this issue is especially relevant with large sample sizes, and identify that p-values address the issue of statistical significance while confidence intervals help to assess what may plausibly be the size of the parameter.

Just like handedness, in which 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. 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. Record whether you have left-eye or right-eye dominance.

There are several research papers (see, for example, Ehrenstein et al., 2005) on eye preference that say that the population proportion of right-eyed people is two-thirds. Let's assume for now (although we would not know this when conducting the study) that this population proportion, symbolized by π, is equal to 0.667.

1. Suppose that you take a random sample of 100 students and find the sample proportion of right-eye dominant students. Do you expect that the sample proportion will equal 0.667? Why or why not?

2. Suppose we calculate a confidence interval from this sample proportion. Do you expect that the interval will contain the value 0.667? Why or why not?

3. Suppose that you select another random sample of 100 students. Is there any guarantee that the sample proportion will be the same as for the first sample? Will the confidence interval based on the new sample necessarily be the same as the confidence interval based on the first sample? How do you think they will differ?

4. To explore the behavior of confidence intervals arising from different random samples with a sample size of 100 from the same population with 50% "successes", open the **Simulating Confidence Intervals** applet.
   - Make sure that **Proportions** and **Wald** are selected from the pull-down menus.
   - Set π to be 0.667 and ***n*** to be 100.
   - Keep **Number of intervals** set to 1.
   - Make sure that **Confidence level** is set to 95%.
   - Press **Sample**.

Notice that this adds an interval (depicted by a horizontal line) to the graph in the middle labeled "*Confidence Intervals*" and a dot corresponding to the randomly generated sample proportion value to the graph labeled "*Sample Statistics*." Click on the interval. Doing this reveals the value of the midpoint ($\hat{p}$) (at the bottom) and the endpoints of the 95% confidence interval based on the sample proportion.

Record the following:

a. Sample proportion:

b. Endpoints of the 95% confidence interval

Lower endpoint: Upper endpoint:

c. Is the sample proportion within the confidence interval? Explain why this should not surprise you.

d. Is the specified long-run proportion (π = 0.667) within the confidence interval?

Notice that 0.667 is either in the interval reported in #4b or not. Suppose the interval turned out to be (0.638, 0.695). It's not like 0.667 is sometimes between 0.638 and 0.695 and sometimes not between those two values. Similarly, even if we didn't know the value of π, the long-run proportion is still some value; it doesn't change if you take a different sample. Therefore, it's incorrect to make a statement like "there is a 95% probability that π is between 0.638 and 0.695." This doesn't fit our "long-run proportion" interpretation of *probability*. So, what is happening 95% of the time?

5. Return to the **Simulating Confidence Intervals** applet.

• Change **Number of intervals** from 1 to 99 for a total of 100 random samples from a process where $\pi = 0.667$.

• Press **Sample**.

Notice that this produces a total of 100 random values for the sample proportion, which are graphed in the graph labeled "*Sample statistics (CI midpoints)*" and the 100 confidence intervals (with 95% confidence) that are based on these sample proportions.

a. Study the graph labeled "*Sample statistics (CI midpoints)*" closely. Around what number is this graph of sample proportions centered (the mean)? Were you expecting this value? Why or why not?

Notice that some of the dots on the "*Sample statistics (CI midpoints)*" graph are colored red and some are colored green.

b. Click on any one of the *red* dots. Doing this will reveal the value of the sample proportion (midpoint) and the corresponding 95% confidence interval. Record the following:

i. Sample proportion

ii. Endpoints of the 95% confidence interval

iii. Is the sample proportion within the confidence interval?

iv. Is the probability you set for the overall process ($\pi = 0.667$) within the confidence interval?

c. Now click on any one of the *green* dots. Doing this will reveal the value of the sample proportion and the corresponding 95% confidence interval. Record the following:

i. Sample proportion

ii. Endpoints of the 95% confidence interval

iii. Is the sample proportion within the confidence interval?

iv. Is the long-run proportion of orange ($\pi = 0.667$) within the confidence interval?

The graph in the middle (*Confidence intervals*) displays the 100 different 95% confidence intervals, depicting
in green the ones that succeed in capturing the actual value of the long-run proportion (which, you'll recall, we assumed to be $\pi = 0.667$) and in red the intervals that fail to capture the actual value of π.

d. Now change the Number of intervals to 100 and press the **Sample** button a few times and watch how, if at all, the percentage reported under "*Running total containing* π" changes. Now record what you think the percentage would be if you were to repeat this process forever.

e. Based on your observations from #5b–c, fill in the blanks:

Thus, 95% confidence means that if we repeatedly sampled from a process and used the sample statistic to construct a 95% confidence interval, in the long run, roughly _______% of all those intervals would manage to capture the actual value of the long-run proportion, and the remaining _______% would not.

6. Now consider changing the confidence level to 90%.

a. Before you make this change, first predict what will happen to the resulting confidence intervals:

i. How will the widths of the confidence intervals change (if at all)?

ii. How will the breakdown of successful/unsuccessful (green/red) intervals change? (What percentage of intervals do you suspect will be green in the long run?)

b. Now change **Confidence level** to 90% and (watch what happens to the intervals as you) press **Recalculate,**
i. How do the widths of the intervals change? Why does this make sense?
ii. How does the running total change? Why does this make sense?

Key Idea
The ***confidence level*** indicates the long-run percentage of confidence intervals that would succeed in capturing the (unknown) value of the parameter if random samples were to be taken repeatedly from the population/process and a confidence interval produced from each sample. So, a 95% confidence level means that 95% of all samples would produce an interval that succeeded in capturing the unknown value of the parameter.

c. Now consider making the sample size four times as large, that is, 400 students, while keeping the confidence level at 90%. Before you make this change, first predict what will happen to the resulting confidence intervals:
i. How will the widths of the confidence intervals change (if at all)?
ii. How will the percentage of successful (green) intervals change (if at all)?

d. Now change ***n*** to 400 and watch what happens as you press **Sample**.
i. How do the widths of the intervals change compared to those from the smaller sample size? Why does this make sense?
ii. How does the running total change? Why does this make sense?

e. To check your understanding about what 99% confidence means, suppose that you calculate a 99% confidence interval for the long-run proportion of right-eyed students to be (0.628, 0.747). For each of the following statements, indicate whether or not it is valid. If you think it is invalid, explain why.
i. There is a 99% chance that the long-run proportion of students who are right-eyed is between 0.628 and 0.747.
ii. We are 99% confident that the long-run proportion of students who are right-eyed is between 0.642 and 0.733.
iii. If we were to repeat the process of taking random samples of 100 students and then calculate a 99% confidence interval from the sample data for each of those samples, then in the long run, 99% of all those confidence intervals would contain the long-run proportion of students who are right-eyed.

References:
Graefe's Arch Clin Exp Ophthalmol (2005) 243: 926–932 DOI 10.1007/s00417-005-1128-7 "Eye preference within the context of binocular functions"

Additional References:
Article about how eye dominance may change:
https://www.sciencedirect.com/science/article/pii/S0960982201004961
Meta-analysis of relationship between handedness and eye dominance:
https://www.tandfonline.com/doi/abs/10.1080/713754206