Exploration 3.1

Malaria in Yemen

Statistical Inference: Confidence intervals

LEARNING GOALS

- Complete multiple two-sided tests of significance, using the same value for the sample proportion but changing the value under the null, and obtain an interval of plausible values for the population parameter.
- Interpret an interval of plausible values as estimating the population parameter and as a confidence interval.
- Based on the results of a test of significance, infer whether or not a value is in the confidence interval and vice versa.

"Statistical inference" is the process of using data from a sample and making conclusions about the population based on that sample. There are two primary types of inference – significance and confidence. Statistical significance "tests" specific values of a parameter as plausible or not (e.g., would I be surprised to see such a result for the sample statistic if the hypothesized value of the population parameter was true?). Statistical confidence estimates the value of the population parameter with an interval of values (e.g., what are the plausible values of the parameter that could have produced this sample?)

In this exploration, we will see how we can use a ***confidence interval*** to estimate a long-run proportion (probability) or population proportion. In particular, we will see how one way to create a confidence interval is to consider many different significance tests about the parameter.

In countries like Yemen, the population has been dealing with endemics such as malaria long before the COVID-19 pandemic hit the world. To estimate the proportion of Malaria cases, a cross-sectional study was conducted on febrile patients (patients presenting with a fever) from November 2018 to April 2019. Patients included were from three districts of Hodeidah City, the second largest city in Yemen, who were referred to the laboratories of the hospitals; 355 volunteered to participate. Do you think these participants are representative of all people in Yemen? How would you estimate the proportion of people in Yemen who have Malaria?

1. Identify the population and sample in this study.
2. Identify the variable recorded in this study. Classify the variable as categorical or quantitative.

Use the symbol π to denote the proportion of all febrile patients in the districts of Hodeidah City where the data were collected who would test positive for malaria.

3. Is π a parameter or a statistic? Explain how you are deciding.
4. Do we know the exact value of π based on the observed data? Explain.
5. State the appropriate null and alternative hypotheses, both in words and in terms of the parameter π, for testing the conjecture that febrile patients presenting in Hodeidah City, Yemen are just as likely to be diagnosed with malaria as not.

Of the 355 patients observed, 115 were confirmed cases of Malaria.

6. Calculate the sample proportion of the observed patients who had Malaria. Also indicate the symbol used to denote this value.

To test whether it is plausible that the sample proportion came from a population where 50% had Malaria, we can generate many random samples of 355 patients, and see how unusual it is for 32.4% of them to have Malaria. A two-sided p-value looks for results at least as extreme as 0.324 in either direction.

7. Conduct a simulation analysis (using the **One Proportion** applet) to assess the strength of evidence that the sample data provide for the conjecture that febrile patients are equally likely to be diagnosed with Malaria as not. Report the approximate two-sided p-value and summarize your conclusion about this strength of evidence.

Your simulation analysis should convince you that the sample data provide very strong evidence to believe that febrile patients are diagnosed with Malaria at a rate different from 0.50 in the long run. That leads to a natural follow-up question: What is the probability of a Malaria diagnosis in Yemen? In other words, we have strong evidence that the long-run proportion of patients presenting with Malaria is different than 1/2, but can we now estimate the value for that probability? We will do this by testing many different (null) values for the probability that a febrile patient is diagnosed with Malaria.

8. Now test whether the data provide evidence that the probability that a febrile patient is diagnosed with Malaria (π) is different from 1/3 (0.333). Use the **One Proportion** applet to determine the two-sided p-value for testing the null value of 0.333. Report what you changed in the applet and report your p-value.

We can specify a ***level of significance*** in order to decide whether the p-value is small. For example, we can say a p-value of 0.05 or less is strong evidence against the null hypothesis and in favor of the alternative hypothesis.

Definition
The ***significance level*** is a value used as a criterion for deciding how small a p-value needs to be to provide convincing evidence to reject the null hypothesis.

Thus, we can *reject* the null hypothesis when the p-value is less than or equal to 0.05. Otherwise, when the p-value is greater than 0.05, we do not have strong enough evidence against the null hypothesis and so we consider the null value to be *plausible* for the parameter.

Key Idea
We will consider a value of the parameter to be plausible if the two-sided p-value for testing that parameter value is larger than the level of significance.

9. Is the p-value for testing the null value of 0.333 less than 0.05? Can the value 0.333 be rejected, or is the value 0.333 plausible for the probability that a febrile patient in Yemen will be diagnosed with Malaria?

The p-value you found in #9 should not have been smaller than 0.05. Hence, you do not reject the null hypothesis at the 0.05 level of significance and therefore you do not reject 0.333 as a plausible value for

π. Thus, it is plausible (i.e., believable) that the probability a febrile patient in Yemen will be diagnosed with Malaria is 0.333.

10. Does this mean that you've *proven* that exactly 33.3% of febrile patients have Malaria? Why or why not?

Because there are still other plausible values, now we want to "zoom in" on which values for the probability are plausible and which can be rejected at the 0.05 significance level.

11. Use the applet to test the probability values given in the following table.

- Each time, change the **Probability of success** to match the value that you are testing (keeping the observed sample proportion that you count beyond the same).
- Everything else should stay the same; press **Draw Samples** and then **Count** to see the new two-sided p-value (with the **Two-sided** box checked). You can also check the **Show sliders** box and double click on the orange value to change the value of π and press Enter and/or use the box and arrow keys to move the slider left and right

Probability under Ho	0.263	0.273	0.283	0.293	0.303	0.313	0.323	0.333
(Two-sided) p-value								
Reject or plausible?								

Probability under Ho	0.343	0.353	0.363	0.373	0.383	0.393	0.403
(Two-sided) p-value							
Reject or plausible?							

12. Using a 0.05 significance level and your results from #11, provide a list of plausible values for π, the long-run proportion of febrile patients in Yemen who are diagnosed with Malaria.

This list of values represents an interval containing all values between two *endpoints*.

Key Idea

The interval of *plausible* values is also called a ***95% confidence interval for π***. Why is this called *95% confidence*? This corresponds to the 5% (0.05) significance level that was used to decide whether there was enough evidence against a hypothesized value. Notice that 95% and 5% add up to give 100%. The 95% is called the ***confidence level***, and is a measure of how confident we are about our interval estimate of the parameter. In this case, we are 95% confident that the long-run proportion of times a febrile patient in Yemen is diagnosed with Malaria is in the interval that we found.

13. Does your 95% confidence interval from #12 include the value of 0.50? Does it include the value 0.333? Explain how your answers relate to the significance tests and p-values that you calculated in #7 and #8.

We can interpret our confidence interval as "I am 95% confidence that the population (long run) proportion of febrile patients in Yemen who have Malaria is between the lower and upper bounds of the confidence interval just constructed." Notice the interpretation of the interval is concerning the population proportion, not the sample proportion. The sample proportion is an estimate of the population proportion and is at the center of the interval. Also notice that we don't say "95% of the time the population proportion is between the lower and upper bounds of the interval." This makes it seem like the interval is fixed and the parameter is bouncing around. This is not the case as the population proportion is fixed. The intervals are what varies with each new sample.

14. Now suppose we were to use a significance level of 0.01 instead of 0.05 to decide whether or not to reject the corresponding null hypothesis for each listed value of π. How would you expect the interval of plausible values to change: wider, narrower, or no change? Explain your reasoning.
15. Implement the 0.01 significance level to determine plausible values for the long-run proportion of times a febrile patient in Yemen is diagnosed with Malaria. (*Hint:* Start with the null values listed in the table in #11, although you might have to test more null values as well.) Report the interval of plausible values.
16. How confident are you about the interval of plausible values that you listed in #15?
17. What is the primary difference between the 95% confidence interval reported in #12 and the 99% confidence interval reported in #15? Which interval is wider? Is this consistent with what you predicted in #14?
18. How would you expect a 90% confidence interval to compare to the 95% and 99% confidence intervals? Explain.

Key Idea
As the confidence level increases, the *width* of the confidence interval also increases.

You should have found very strong evidence, based on a small p-value, that a population proportion different from 0.50 of febrile patients in Yemen are diagnosed with Malaria. Moreover, your confidence intervals, which provide plausible values of the population proportion diagnosed with Malaria, contain only values different from 0.50, confirming the conclusion made from the p-value.

In fact, confidence intervals and tests of significance generally give complementary information. Values that are not considered plausible by the (two-sided) test of significance should not be contained in the corresponding confidence interval (*confidence level* = 100% – *level of significance*). Because 0.50 is not inside the confidence interval, this matches the test of significance. The confidence interval not containing the hypothesized value is not "further evidence against the null" but confirms that your calculations agree. The confidence interval then gives us additional information about the size of the prevalence (long-run proportion) of Malaria.

Rashad Abdul-Ghani ,Mohammed A. K. Mahdy, Sameer Alkubati, Abdullah A. Al-Mikhlafy, Abdullah Alhariri, Mrinalini Das, Kapilkumar Dave, Julita Gil-Cuesta. Malaria and dengue in Hodeidah city, Yemen: High proportion of febrile outpatients with dengue or malaria, but low proportion co-infected

Published: June 25, 2021
https://doi.org/10.1371/journal.pone.0253556