

# Introduction to Statistical Inference

## Hospital Visits and Admissions

Me, Analyst

Oct 14, 2025

### About This Activity

In this activity, we'll examine data about hospital visits that result in admissions and patient refusals at a particular hospital. We'll treat the background data set as the entire population, and use sampling methods for estimating or making claims about patient admission rates.

**Task 1:** Run the following code cell to load the `{tidyverse}` ecosystem of packages and read in our `hospital_visits` data.

**Task 2:** How many patient visits are included in the `hospital_visits` data set? What are the variables included?

**Task 3:** Without looking at any of the data, conjecture and justify an admission rate. Why is it difficult to do this?

**Task 4:** Run the code below to take a small sample of patient visits and calculate the admission rate for the hospital. Before you proceed, describe what each line of code is doing.

**Task 5:** Revisit your conjectured admission rate from *Task 3*, do you have more, less, or the same level of confidence in your initial estimate? Why?

**Task 6:** Did you get the same result as your neighbors? Discuss why this is the case. Uncomment the code `set.seed(241)` and revisit this question? What do you think `set.seed(241)` does?

**Task 7:** What are some ways we could improve our estimate for the patient admission rate at this hospital? Are some methods better than others?

**Task 8:** Run the code cell below to draw a sample and construct and plot the corresponding confidence interval for the proportion of patients admitted to the hospital.

**Task 9:** Run the code cell above several times in order to see what happens. Describe the result.

**Task 10:** What likely prohibits us from being able to build multiple confidence intervals like we did above in real applications?

**Task 11:** Run the code cell below to investigate what is meant by an XX% confidence interval. We'll draw 100 random samples and then construct and plot those confidence intervals.

**Task 12:** Knowing that we're going to obtain one confidence interval, not 100 of them, how should we interpret the interval we obtain?

**Task 13:** How do the confidence intervals change when you alter the `confidence_level`? What about if you change the `sample_size`?

**Task 14:** Suppose now that we focus only on patient visits where the primary complaint was flu-like symptoms. Should we use a confidence interval that we calculated above (for all patients) as a tool for capturing the proportion of influenza-prompted visits resulting in a hospital admission? Discuss why or why not.

**Task 15:** Suppose the hospital claims that their true admission rate resulting from patients complaining of flu symptoms is over 0.20 (that is, more than 20% of patients who visit the hospital with complaints of flu symptoms are admitted). Is it possible to test the hospital's claim? Discuss how you might go about testing the claim. What kind of evidence would support this claim? What kind of evidence would contradict it?

**Task 16:** The code cell below takes a random sample of patients complaining of flu symptoms only and calculates the proportion of these visits resulting in a hospital admission. How does your sample estimate compare to the hospital's claim of 0.20? Would you say your result provides strong, weak, or no evidence against the claim? Would your conclusion necessarily be the same if you drew a different sample?



#### What We've Learned

- Confidence intervals estimate a plausible range of values for a population parameter.
- Confidence levels describe long-run success rates of such intervals.
- Sampling variability means different samples produce different estimates.
- When someone makes a claim about a parameter, we can use data to evaluate whether that claim seems consistent with what we observe.