

# **01 Introduction to Different Statistical Paradigms**

**Comparing Frequentist and Bayesian Approaches**

Stefani Langehennig      Zach del Rosario

## Introduction

This activity is all about *statistical inference* and *statistical assumptions*. Both are important when you have data and want to use that data to help you better understand some phenomenon about the world. For example, in [00 Sustainability Trends Among Disadvantaged Communities](#), we used the CEJST dataset to understand general trends around sustainability and disadvantaged communities. However, at this point, we have not crafted specific research questions or hypotheses that we would like to rigorously test with the data we have. Nor do we know, with any degree of certainty, how “correct” the trends are that we saw in our data or all the ways that this trend could have occurred.

There are a few different ways that we can test our hypotheses using methods of statistical inference. Statistical inference allows us to take some data, create a model, and make sense of the complex world around us. Each method has a set of assumptions built into it, which influences the approach you take to answer your research questions, as well as the way you interpret your findings. In the activity that follows, you will have the opportunity to use methods of statistical inference to answer a research question and explain, based on the set of assumptions baked into your approach, your ultimate findings.

## Learning Objectives

By the end of the activity, you should be able to:

1. Evaluate multiple hypotheses using inferential statistical results
2. Connect your evaluation of hypotheses with real-world factors
3. State the primary statistical assumptions for Frequentist and Bayesian inference, and understand how they can lead to different conclusions

### Instructor Notes: Introduction to Statistical Paradigms

By the end of the activity, you should be able to:

1. Evaluate multiple hypotheses using inferential statistical results.
  - The hypotheses we consider in this activity concern patterns of inequality: Do Black Americans experience a disproportionate level of energy expenditure?
  - The inference we consider is across U.S. States: Do patterns of inequality persist across states? What do we conclude when we have limited data?
2. Connect your evaluation of hypotheses with real-world factors.
  - We ask students to think critically about *why* patterns of inequality do—or do not—appear in different states. E.g., what is different about policy in Colorado vs Florida?

3. State the primary statistical assumptions for Frequentist and Bayesian inference, and understand how they can lead to different conclusions.

- We will focus on the inclusion of a prior and its interaction with small/large sample size. A prior is particularly useful when we have limited data. The prior becomes less important when we have a large dataset.

## Before We Begin...

Ask yourself the following questions:

### Warmup Questions

- What do you remember about analyzing a dataset with statistics?
- What do you know about frequentist and Bayesian statistics? (*Note: It's okay if you don't know what this means!*)

Discuss these questions with the people around you.

## The Big Idea: Inference

Statistical inference is drawing conclusions about **data we haven't seen** using **only the data that are available**. As a simple example, if we're interested in the energy burden pattern in Colorado but we only have data from Massachusetts, we should be careful when trying to use the MA data to predict patterns in CO. Crucially, **the assumptions we make for inference will strongly affect the conclusions we make**.

All models are wrong....

All statistical analyses involve making assumptions—we'll practice making and assessing statistical assumptions in this activity.

## Crafting Research Questions & Hypotheses

Ahead of using data and statistical inference, we usually create theory-based *research questions* and *hypotheses* that help guide what we expect to find. Both are focused on an aspect of the world that we want to know more about. However, research questions tend to be open-ended, allowing for discussion on what the outcome might be and what might explain it. For example, a research question about sustainability and diverse communities may look like this:

*Why does the projected flood risk vary among different communities in U.S. Census tracts?*

Hypotheses, on the other hand, tend to be closed-ended, articulating a certain relationship between an outcome and the things that influence that outcome. For example, one of the testable hypotheses stemming from the research question on projected flood risk may be:

*As the population density of Hispanic communities increases (decreases), the projected flood risk increases (decreases).*

Note that this hypothesis is succinct and testable with the data we have. It is also directly related to our more general research question.

## Paradigms in Statistical Inference

There are a few different ways to draw conclusions about questions we have using data. We will focus on two inferential perspectives in this class: *Frequentist* and *Bayesian*. For this activity, we will discuss these differences in the context of **general inference** and **model summaries**.

For general inference, we are concerned with answering questions about a *larger* dataset, while only having access to a *smaller* dataset. We do this by:

1. Translating our question into a mathematical model, which requires assumptions
2. Fitting the model using data
3. Interpreting model results in terms of our question

After we have done the general inference steps above, we must think about how to translate what we have done and what we have found.

- Statistical models are complex objects, so we use *model summaries* to help.
- Statistical models represent the *uncertainty due to limited data*.

Some common summaries we use:

Table 1: Common Model Summaries

Point estimate	Uncertainty	Interval
42	4	$(42 \pm 2 \times 4) = (34, 50)$

Stay Tuned!

Note that we haven't talked about frequentist or Bayesian statistics yet! We'll get there later in the activity....

You can read more about the statistical inference in [03 One Page Summary](#).

## Next Steps

Given the context we have on sustainability trends among disadvantaged communities using just a few of the variables in the CEJST dataset, we can develop a research question and hypotheses to test using methods of statistical inference:

### Research Question:

*Do Black Americans experience a disproportionate level of energy expenditure?*

### Hypothesis:

*As the population of Black Americans increases (decreases), the level of energy expenditure increases (decreases).*

**Coming up next:** When you get here, please discuss the reading with your tablemates. When you all feel that you (at least somewhat) understand the reading, please start working through 02 CEJST Activity with the other students at your table.

We are going to test our hypothesis and answer our research question using the CEJST dataset.