04 Activity Wrap-Up

Comparing Frequentist and Bayesian Approaches

Stefani Langehennig Zach del Rosario

Overview

In this activity, we learned more about *statistical inference* and *statistical assumptions* through the lenses of Frequentists and Bayesians. We used the CEJST dataset to understand general trends around sustainability and disadvantaged communities, as well as crafted specific research questions and hypotheses that we would like to rigorously test with the data we have. Our research question asked:

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

And our primary hypothesis we wanted to test is:

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

Let's discuss what we've found as a class:

Questions for the Class

- What can we say about our hypothesis? Does it depend on U.S. State?
- How would you answer our research question now that we have analyzed the data?
- What can we conclude about the relationship between sustainability and disadvantaged communities? What might you recommend from a policy-making perspective?
- **Can** we make sound policy decisions, based on what we've seen so far?

Instructor Tips

- Deliberately call on students from frequentist **and** Bayesian groups.
- Try to highlight the differences in the students' findings:
  - "Did anyone find a negative trend for Colorado? Did anyone find a positive trend?"
  - "Why might these be different?"
- Emphasize that you can't start making sound policy decisions until you understand discrepancies in your analysis.

NOTE: Please do not go on to the next page until your instructor tells you to!

Differences between Frequentist & Bayesian Approaches

There are key differences between Frequentist and Bayesian approaches. In the activity, we focused on model summaries and general inference.

General Inference

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

1. Translate our question into a mathematical model, which requires assumptions
2. Fit the model using data
3. Interpret model results in terms of our question

Table 1: General Inference Comparison

Frequentist	Bayesian
Provides: Pr(data | hypothesis) Decision based on hard cut-off	Provides: Pr(hypothesis | data) Decision based on posterior distribution

Model Summaries

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

Some common summaries:

Table 2: Common Model Summaries

Point estimate	Uncertainty	Interval
42	4	(42 +/- 2x4) = (34, 50)

Here are the differences between how Frequentists and Bayesians get model summaries:

Table 3: Model Summary Comparison

Frequentist	Bayesian
Maximum likelihood estimate	Maximum of the posterior density
Standard error	Posterior standard deviation
Confidence interval	Credible interval

Applying the Differences

With the remainder of class time, separate into groups of 2-3 people, where each group should have at least 1 person who did the Frequentist analysis and 1 person who did the Bayesian analysis. Using the critical differences listed on the previous page, discuss the following:

- How do our modeling choices affect our outcomes of interest based on the assumptions we know about both approaches? Are the results for our hypothesis different? If so, how?
- How do we interpret the coefficients (m and b) in both the Frequentist and Bayesian models?
- How might we improve our model based on what we know from our **model summaries** and **general inference** for both approaches?

Jot down 2-3 sentences for each question, as well as any remaining questions or concerns you have about conducting your analyses using one of the two approaches.

Instructor Notes: Activity Wrap-Up

While the conclusion document provides a fairly structured approach for closing out the activity, we anticipate this portion of the activity will be organic depending on how the instructor taught the materials and how the students received them. Simply, instructors are responsible for engineering the "controversy" in the classroom so that the open disagreement between students and their respective approaches is maximized as part of the learning experience.

- At the end of the discussion, if time, have the groups share what they came up with based on their answers.
- This should be a lively debate facilitated by the instructor; students will debate who is "correct"
- Instructor should highlight the differences between their analyses.
  - All students started with an identical dataset and framing questions; it is the difference in their starting assumptions (i.e., a frequentist or Bayesian approach) that results in different conclusions.
  - Instructor should also consult the 5E Model Approach outlined in the `run-of-show.md` document to help facilitate this discussion
  - Note that the frequentist and Bayesian approaches will tend to converge to the same results in the limit of more data; you can use the full USA results to illustrate this point.
  - Takeaway point: controversy isn't resolved, however statistical results are conditional on the assumptions chosen for the analysis and implies that choosing

appropriate assumptions is critical to a sound analysis

Coming up next: When the instructor tells you, please flip to the last page of this activity and follow the QR code link to a post-activity survey. **Please do not move on until your instructor tells you to.**

NOTE: Please do not flip to the last page until your instructor tells you to.

Please use this QR code to access the post-activity survey.