Surprise - They're Different!

Comparing Frequentist and Bayesian Approaches in Public Policy

Stefani Langehennig, PhD[1,*] Zach del Rosario, PhD[2]
Mine Dogucu, PhD[3]

Implementing quantitative methodological techniques is a crucial piece of understanding public policy. While quasi-experimental, spatial (diffusion), and mixed methods are most commonly used when teaching policy studies, little research exists on using Bayesian approaches for policy learning, or how the outcomes from traditional quantitative approaches differ from a Bayesian approach. We propose an applied learning activity for students of public policy that exposes them to Bayesian methods and explores the differences between this statistical paradigm and more commonly used approaches. We do this using a structured interrogation of the The Climate and Economic Justice Screening Tool (CEJST) and the epistemological framings in the 5E model (Elby and Hammer 2010; Duran and Duran 2004). The activity illustrates the importance of statistical assumptions, and by extension, the impact that different quantitative methods have on understanding public policy. The goal of the study is to introduce students, instructors, and practitioners of public policy to a new way of using statistics, equipping them with the tool set and logical processes necessary to apply either approach as they see fit when studying public policy.

[1] Assistant Professor of the Practice, University of Denver
[2] Assistant Professor, Olin College of Engineering
[3] Associate Professor, University of California Irvine

[*] Correspondence: Stefani Langehennig, PhD <stefani.langehennig@du.edu>

Working paper: please do not cite without the permission of the authors.

Introduction

Social scientists routinely make use of quantitative methods to understand the complex world around them. Approaches employed range from quasi-experimental, spatial (diffusion), and econometric techniques, to methods that are more qualitative in nature. While Bayesian approaches are not completely missing from public policy and related fields (see Gill and Witko 2013; Fienberg 2011), they are underutilized for policy learning among academics and policy specialists. This omission is for a few reasons. First, scholars and practitioners in government and public policy spaces often are not exposed to Bayesian methods when they are taught quantitative techniques (Gill and Witko 2013). Thus, the lack of familiarity means Bayesian techniques are not a regular tool in their methods toolkit. Second, the reliance on infusing subjective priors into the methodological approach tends to be questioned by non-statisticians who have received more traditional Frequentist training in their disciplines (Fienberg 2011; Freedman 1997).

In this paper, we propose an applied learning activity for students of public policy - though its applications extend beyond this discipline - that exposes them to Bayesian methods, focusing on both the theoretical and practical underpinnings of the approach. The activity explores the differences between Bayesian approaches and more commonly used statistical techniques to introduce students, instructors, and practitioners to new and different ways of using statistics to investigate real-world problems. Through the application of this activity, our hope is that using these methods will equip them with the tool set and logical processes necessary to apply different quantitative approaches as they see fit when studying public policy. More importantly, using this activity in social science classrooms will encourage students to interrogate and challenge differences in the assumptions made, and outcomes produced, in the statistical approaches they are taught.

What follows is an in-depth examination of how teaching different quantitative methods results in a more robust understanding of public policy. We start by reviewing the research to date on cross-disciplinary approaches to using quantitative methods, as well as how this influences students' assumptions about their own learning using the epistemological framings in the 5E Model (Elby and Hammer 2010; Duran and Duran 2004). Next, we introduce our applied activity and describe its implementation in the classroom. We then discuss our findings from implementing the activity, followed by our conclusions from the study.

Statistical Inference for Non-Statistics Disciplines

Conveying complex concepts and analytical techniques is one of the most challenging aspects of teaching (Bates and Jenkins 2007). In addition to distilling topics down to make them accessible to all students, instructors are frequently tasked with ensuring that the topics learned in class translate "out in the wild" when students enter their respective professions. This is true in the social sciences and other non-statistics disciplines, like public policy, where students are tasked with translating what they know into evidence-based practices. While the theories underpinning non-statistics subjects can be challenging to understand, layering the quantitative methods typically needed to test research questions on top of these theories can create added challenges. Too often the methods and theories, alongside the question of usefulness outside of the academy, make it difficult to link analysis and policy practices together (Connelly et al. 2021).

What is more, students who are not statistics majors frequently become anxious and tend to avoid quantitative methods unless the methods are contextualized and the connection between their subject and technical approach seems clear (Gunn 2017). Nonetheless, quantitative methods - particularly statistical inference - are a must for social science students. Most students take at least a one to three class sequence on quantitative methods as undergraduates. These usually include some form of research methods, followed by an applied class that teaches traditional Frequentist statistical inference (for example, Null Hypothesis Significance Testing (NHST)).

Though they are commonly overlooked relative to other forms of statistical inference, Bayesian approaches do make an appearance in the social sciences. For example, Jackman (2000a) offers Bayesian simulation using Markov chain Monte Carlo (MCMC) algorithms as a unifying quantitative solution to estimating models across several contexts in political science (see also Jackman 2000b, 2004, 2009). Shor et al. (2007) compare simulations of Bayesian multilevel models to other standard estimators for hierarchical data to showcase the Bayesian model's flexibility and robustness in time-series cross-sectional data, which are commonly used in political science. Still others use in-person interviews to form elicited priors to better understand trust in judicial systems, bridging the gap between qualitative information and empirical models (Gill and Walker 2005). Even in the realm of law, scholars are utilizing techniques such as Bayes factors to buttress forensic evidence in expert testimony in court cases (Kadane and Nordgaard 2024).

In public policy, Bayesian techniques have been used to analyze the most critical of events, including the recent Covid-19 pandemic. For example, Wibbens, Koo, and McGahan (2020) use Bayesian inference to investigate the effectiveness of Covid-19 control strategies across the world, finding broadly executed "core policies" (staying home, cancelling events, etc.) and compliant jurisdictions somewhat reduced the spread of the virus. Deslatte, Tavares, and Feiock (2018) use Bayesian multilevel models to show that policymakers in city government use permit delays for land use as a tactic to manage growth. Practitioners of public policy are turning to Bayesian inference to tackle the challenge of interpreting findings from impact evaluations, creating new frameworks, such as BASIE (BAyeSian Interpretation of Estimates) (Deke, Finucane, and Thal 2022), to "improve evidenced-based decision making."[1]

Given academics and practitioners outside of statistics use Bayesian techniques for their own work, a handful have tried to incorporate Bayesian philosophy and principles into how estimation is taught among their students. For instance, Gill and Witko (2013) offer an accessible introduction to Bayesian analysis for students of public administration and policy, arguing that understanding the theoretical differences between the two approaches is critical, and that Bayesian techniques are more appropriate for the discipline. Wagner and Gill (2005) make a similar argument, stressing that Bayesian inference is "better suited" for answering public policy questions, showcasing their application on educational outcomes in

[1] https://www.mathematica.org/features/bayesian-methods

public schools. Indeed, they even go so far as to say that the traditional statistical inference methods taught in public administration and policy programs is "defective and should be replaced (Wagner and Gill 2005, 5)" because they are borrowed from other disciplines and do not necessarily fit their research challenges. Finally, Luque and Sosa (2023) use a series of different scenarios to demonstrate ways in which to incorporate Bayesian methods into real-world social inquiries. They do this by way of inferential comparisons between Frequentists and Bayesians, underscoring the differences between the two approaches both theoretically and in practice.

Activating Epistemological Frames

One of the primary points stressed in the research advocating for Bayesian inference in non-statistics disciplines is not to simply introduce another applied tool without much thought; rather, offering different philosophies and theoretical approaches in addition to the tool is crucial to employing effective analytical practices. Simply, being able to choose and understand the impact of the methodologies used when approaching a real-world problem is critical. In social science quantitative methodology teachings, intellectual development often starts with a "surface approach", where students memorize and repeat facts that inform them, rather than engaging with or reflecting on the approaches they have learned (Entwistle 1997; Bates and Jenkins 2007). Our goal, in addition to offering a Bayesian learning activity, is to move beyond the surface approach, empowering students with the knowledge to choose between different statistical approaches based on what is productive for the context at hand. This starts with epistemological framings, or the nature of knowledge and understanding among students (Elby and Hammer 2010).

At its most basic level, epistemology is concerned with the methods and theories that help understand knowledge origins and acquisition. Epistemological framings are created from personal cognitive resources, such as beliefs, that are activated among individuals conditional on the context they are in (Elby and Hammer 2010, 3). These cognitive resources provide a framework with which individuals can form their understanding of the "nature of knowledge, knowing, and learning", and are often variable and something that the person may not be aware of (Elby and Hammer 2010, 4). An epistemological frame is when individual resources that are reinforced by one another (e.g., "locally coherent") are activated and lead to knowledge stability (Elby and Hammer 2010, 6). Simply, these frames are the expectations that someone brings to different experiences and scenarios, which in turn influences their actions.

In the context of the classroom, epistemological frames organize activities in the classroom for students, which in turn affects their knowledge and learning. Students who can recognize and critique the assumptions underpinning their analyses (treating them as tentative), but carry out their analyses respecting those analyses (treating them as true) will likely be more effective as practicing statisticians in their discipline. Getting students to recognize the importance of assumptions—and to practice adopting different assumptions—will be a critical first step in developing these multiple epistemological framings.

Our goal is to activate students' epistemological frames by structuring our activity around the 5E Model proposed by Duran and Duran (2004). The 5E Model is based on constructivist pedagogies and rests on the notion of inquiry-based teaching, or having students discover information on their own without the direct help of instructors (Duran and Duran 2004; Uno 1999). The 5E model is captured in the following figure:

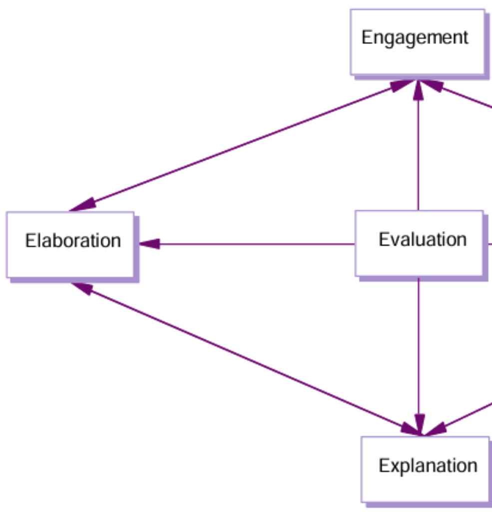


Figure 1: The 5E Instructional Model (as seen in Duran and Duran (2004), p.52)

We discuss the 5E model more below, however each of the E's can be thought of as such: *Engage*: Get students interested; *Explore*: Students do self-directed inquiry; *Explain*: Give students conceptual tools; *Elaborate*: Let students work with the tools; and *Evaluate*: Assess the learning outcomes. We turn next to the applied activity, where students compare Frequentist and Bayesian approaches.

Applied Activity: Comparing Frequentist & Bayesian Approaches

To alleviate some of the challenges that come with teaching different techniques of statistical inference, scholars have compared Frequentist and Bayesian approaches to one another, highlighting where they are similar and when they diverge (Samaniego 2010). We designed, implemented, and evaluated an activity that does just this, using a public policy example to guide our comparison. Our primary work included designing an activity that takes students through a structured interrogation of a dataset, with the pilot activity launching during spring of 2024. The activity is designed to take place during one or two class sessions. We evaluated this activity's impact on students' statistical knowledge and epistemological framings using the 5E Model (Elby and Hammer 2010), discussed more below.

The activity is focused on a real dataset, to which groups of students are given guided statistical analysis. Students follow a structured process to analyze the dataset and interpret their results. However, different groups receive different versions of the activity: some receive a Frequentist approach, while the others receive a Bayesian approach. By carefully crafting the analyses to reach different conclusions, we aim to surprise students with diverging conclusions. The activity concludes with a final full-group discussion, where we highlight the importance of statistical assumptions, completing the comparison of Frequentist and Bayesian approaches.

The activity learning objectives are three-fold. First, students should be able to evaluate multiple hypotheses using inferential statistics; second, students should be able to connect their evaluation of hypotheses to real-world factors; and third, students should be able to state the primary statistical assumptions for Frequentist and Bayesian inference, and understand how they can lead to different conclusions. These learning objectives stem from our overall learning goal of engineering a "classroom controversy" to motivate students to find their own understanding of how Frequentist and Bayesian assumptions can lead to different conclusions (and by extension, real-world decisionmaking).

Recruitment

The work for this manuscript was completed under an IRB exempt protocol (number 2156277) approved by the University of Denver IRB. It is designed for use in a classroom for non-statistics majors. To pilot the activity among students before disseminating widely, we implemented the activity with two distinct populations at our respective institutions. Students from the University of Denver (DU) Daniels College of Business who participated in the activity are undergraduates - typically freshmen and sophomores - who have various majors ranging from Marketing to Business Ethics and Legal Studies. Students recruited from Olin College of Engineering - a small liberal arts-flavored university in Massachusetts - are undergraduates who have focus on a variety of engineering-related topics. After completion of the pilot study, we plan to disseminate and implement this activity with other non-statistics undergraduates, including students of public policy and political science.

For both institutions, participants had to be at least 18 years of age or older and must have completed at least one entry-level statistics course. At the University of Denver, all students enrolled in a business school major must complete three courses that are part of their statistics sequence (INFO 1010, 1020, and 2020). Here, the activity was introduced to students currently enrolled in a Winter 2024 quarter INFO 1020 course. The activity was run during normal class time as a typical lecture; however, students did not receive a grade for the activity and had the option to opt out if they so chose. Similarly, the activity was run during normal class time at Olin with undergraduate students currently enrolled in the Introduction to Data Science course. Students at Olin also had the choice to participate or opt out if they so chose. Students at both institutions were not compensated for participating in the activity. Both populations were exposed to topics such as visualization, data wrangling, and basics of frequentist statistics (e.g., confidence intervals, hypothesis testing, etc.).

Activity Materials

All materials were created using the programming language `R` and can be rendered in .html or .pdf format for use. The materials are openly available for instructors on our GitHub repository. Important starter documents include the run of show, which outlines at a high level the different steps of the activity, as well as the artifacts used in the activity. Further, the learning objectives and details of the activity are fleshed out in the introduction document. Instructors can watch a video overview of the activity on YouTube, and can use the Makefile available in our repository to compile both student and instructor-facing artifacts (the latter contains solutions and other notes for running the activity).

Activty Approach

To implement the activity, there are four steps, each discussed at length below:

1. Setting the context for the real world problem the class is exploring
2. Introducing the motivation for the activity (statistical approaches) given the context
3. Doing the applied activity
4. Closing out the activity

The activity is built around the aforementioned 5E Model Approach, where two loops of the model are working concurrently. The first loop is focused on the applied learning aspects of the activity, or the application of a statistical approach focused on current issues. For example, for *engage* the goal is to motivate students with current issues around climate and equity. The *explore* stage is the opportunity in which students get to do self-directed inquiry. For this activity, that means investigating the real-world dataset provided to them in small groups. *Explain* gives the students the conceptual tools they need to understand the different statistical approaches. Students will learn the basics of assessing and interpreting a fitted statistical model with the instructor. For *elaborate*, students get to work with the tools, meaning they apply the conceptual tools they learned to the real-world dataset. Finally, *evaluate* involves giving students the opportunity to reflect on their understanding of the concepts and application they just did through an instructor-facilitated class discussion.

The second loop is focused on the conceptual learning aspects of the activity, or the ownership of the results that students discovered. *Engage* is focused on the different questions around the data (context) that should motivate their search for an explanation for outcomes using statistical approaches. *Explore* and *explain* capture the introduction to statistical inference broadly defined, the learning objectives for the activity, and the high-level critical differences in Frequentist and Bayesian approaches. For *elaborate*, students articulate the basic concepts of assessing and interpreting a fitted statistical model and come to conclusions related to the research question and hypotheses. Finally, *evaluate* uses the concepts and what students learned in the application to articulate and refine their understanding of the differences between Frequentist and Bayesian approaches.

Important to this approach is the role of the student. While the instructor is there to facilitate questions and the general cadence of the activity, the onus is on students to work through each of the steps of the activity. This is a fundamental tenant of the 5E model, where inquiry-based learning is key to students discovering information and learning.

Problem Context

Given our interest in teaching students new approaches to examining real-world public policy problems, we start our activity by introducing The Climate and Economic Justice Screening Tool (CEJST). The CEJST is the result of President Biden's Executive Order issued in January 2021. The tool is used to identify and subsequently help communities disadvantaged by the burdens stemming from climate change in government social programs. While the data covers a number of burdens (health, transportation, and workforce development, for example), we focus on the sustainability aspects of the tool for our activity, including climate change, energy, and legacy pollution burdens on communities.

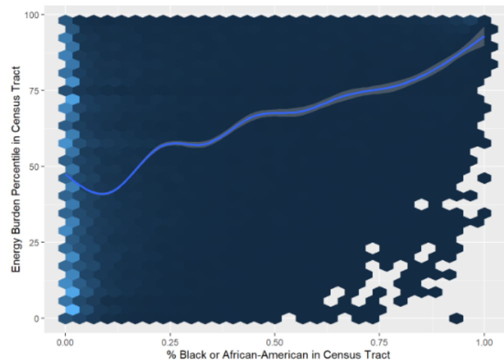
We begin with a straightforward explanation of the dataset, situating it in the contemporary dialogue around climate change. Specifically, we theorize there may be a relationship between climate change burdens and minorities residing in Census tracts. By providing them with this context, we get students to think about possible questions - and by extension, hypotheses - they may be able to explore using statistical inference.

To dig into the context of our real-world problem further, we also provide embedded code snippets and output from `R` of some high-level exploratory data analysis (EDA) for the students to review and discuss. We begin by focusing our attention on a few variables of interest for EDA. We start with the *energy burden percentile*, which captures the percentile of energy cost as well as energy-related pollution within a census tract, as well as the *percent of African-American or Black alone*, which captures the percent of African-American or Black individuals in a census tract.[2] The instructor walks the students through basic data wrangling, providing prompt questions to get them thinking about the substantive implications of descriptive statistics and visualizations. Below is an example of the output, which contains the code used to create the figure, as well as a short description and a prompt question for the students.

[2]While we guide our students to the variables we want to explore for this implementation of the activity, instructors can modify the activity such that students explore the dataset and identify variables of interest on their own. If they know how to use statistical software, they can also conduct the EDA and analyses in `R` on their own, rather than giving them the completed version as we do here.

```
# Plot relationship
p <- df_small %>%
  ggplot(aes(
    percent_black_or_african_american_alone,
    energy_burden_percentile
  )) +
  geom_hex() +
  xlab("% Black or African-American in Census Tract") +
  ylab("Energy Burden Percentile in Census Tract") +
  geom_smooth()

# Remove legend
p + theme(legend.position = "none")
```



As the figure shows, there is a positive relationship between the percent of Black or African-Americans living in a census tract and the energy burden percentile in the respective census tract.

> Visualizing Variables of Interest
>
> What other ways might we choose to visualize our variables of interest? Is there anything else concerning or interesting based on our visualizations?

Figure 2: Example of Context Document

After students have investigated the dataset and have a more thorough understanding of the problem at hand, the instructor turns their attention to the introduction document, discussed next.

Activity Introduction

This portion of the activity is instructor led to begin, walking students through the learning objectives and warmup questions, which in turn initiates a group-wide discussion on statistical inference. The instructor has the students discuss inference at a high level, offering more pointed discussion if needed around crafting a research question and hypotheses. At this point, the instructor turns students to the simplified critical differences one-pager that introduces them to the primary differences between the Frequentist and Bayesian paradigms. To keep the exercise manageable, we focus students' attention on *general inference* and *model summaries*.[3] Table 1 and Table 2 showcase the differences outlined in the critical differences one-pager between general inference and model summaries:

[3]You can access the full one-pager that is instructor-facing on the associated GitHub repository. In addition to comparing general inference and model summaries, it also includes comparisons between fixed variables, interpreting probabilities, and model inference.

Table 1: General Inference

Frequentist	Bayesian
Deduction from Pr(data | H0), by setting α in advance	Induction from Pr(θ | data), starting with Pr(θ)
Accept H1 if Pr(data | H0) $< \alpha$	$1-\alpha$% of most likely parameter values fall within a $1-\alpha$ HPD
Accept H0 if Pr(data | H0) α	

Table 2: Model Summaries

Frequentist	Bayesian
Point estimates and standard errors	Descriptions of the posterior distribution such as means and quantiles
95% confidence intervals indicating that 19/20 times the interval covers the true parameter value	Highest posterior density intervals indicating region of highest posterior probability
	$1-\alpha$% of most likely parameter values fall within a $1-\alpha$ HPD

The introductory discussion of the activity wraps up with the instructor introducing the research question and associated hypothesis the class will test with their respective statistical approach. Specifically, students with will be assessing whether Black Americans experience a disproportionate level of energy expenditure using inferential statistics and the CEJST dataset.[4]

Activity Application

The activity application step is the heart of the exercise. Students are assigned a random number generated by a random number generator and put into groups based on their number to go through an applied statistical analysis. There are two versions that are circulated: the Frequentist analysis and the Bayesian analysis. The analyses that are given to the students are already completed - they only receive the output of the analysis with associated questions to help them think through the different parts of the analysis before they come to any conclusions.

It is important to note that the data used for each analysis is the exact same for both of the groups, as is the hypothesis that the students are testing. Additionally, the students are asked to assess the same conceptual things, regardless of which activity they receive. They

[4] An extension of the activity could have the students develop the research questions and hypotheses on their own. Given the activity we implemented is meant to span one class period, we provide those for the students.

will use the *general inference* and *model summaries* comparison discussed in the Section to diagnose the outputs of the models from the analyses.

Frequentist Analysis

Both activities begin with a quick overview of the hypothesis the students are testing, as well as the different components of a statistical model. For the Frequentist model specifically, the analysis document introduces the following model, where B as the dependent variable (energy burden percentile), P is the percent black, m is the slope parameter, b is the intercept parameter, and ϵ captures the error term.

$$B = mP + b + \epsilon$$

The instructor encourages the class to think through how to interpret estimates in a linear model using Frequentist statistics, noting a number of important assumptions along the way in questions are asked about the model, including that the b and m parameters are fixed but unknown values. This is a natural place for a number of questions to be asked of the class related to model summaries and general inference. The questions below illustrate what the class is asked for understanding model summaries in a Frequentist model, along with the associated answers:

Questions for the Class

- Which scenario gives the largest estimate for the slope?
  - Scenario B
- Does the confidence interval for Scenario A include zero? (NB. A confidence interval includes zero if Lower $< 0 <$ Upper.)
  - Yes
- Does the confidence interval for Scenario B include zero?
  - Yes
- Does the confidence interval for Scenario C include zero?
  - No
- If a confidence interval includes zero, this indicates that we cannot conclude whether the slope is positive or negative. For which scenarios can we not conclude whether the slope is positive or negative?
  - Scenarios A and B

After students have revisited the concepts around model summaries and general inference for the Frequentist linear model, they move to a predictive model where they are given a number of predicted versus observed plots, show below, for the model across three states: Massachusetts, Colorado, Florida, and the entire sample of data (e.g., the United States).

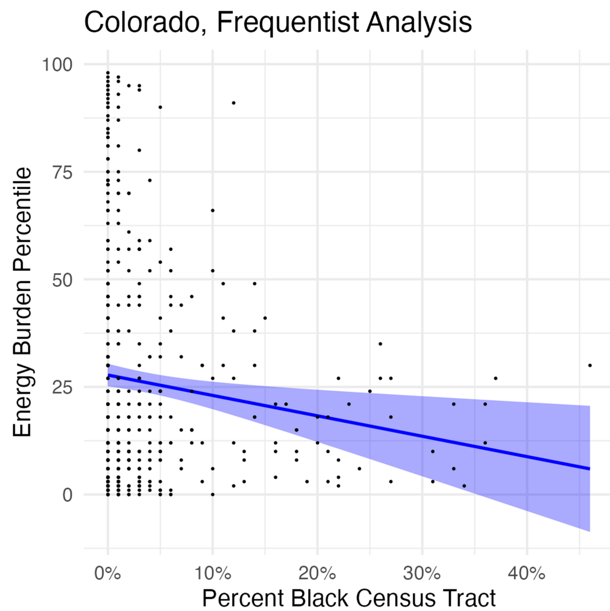
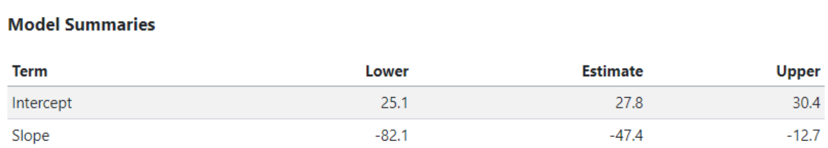


Figure 3: Example Trend from Frequentist Analysis

Additionally, they are given the intercept and slope estimates, as well as the confidence intervals for each model.



Model Summaries

Term	Lower	Estimate	Upper
Intercept	25.1	27.8	30.4
Slope	-82.1	-47.4	-12.7

Figure 4: Example Model Summary from Frequentist Analysis

Once they have seen the results for each of the states and the U.S.(which differ depending on the sample of data being used), students are given a set of questions that encourage them to think about the model results knowing what they do about model summaries and general inference. At this point, the students are exploring the results of the model in their respective groups and discussing and answering together the questions provided to them.

Bayesian Analysis

The cadence of the Bayesian analysis largely mirrors the Frequentist analysis to begin. Like the Frequentist groups, the Bayesian groups also revisit the hypothesis they are testing, as well as the different components of a statistical model, though this time it is a Bayesian linear model.

$$B = mP + b + \epsilon$$

where B is the energy burden percentile, P is the percent Black, m is the slope parameter, b is the intercept parameter, and ϵ is a *residual* term that represents factors not accounted in the model. The residual term is assumed to be normally distributed $\epsilon \sim N(0, \sigma^2)$ with an unknown parameter σ^2. All three parameters have a prior distribution, defined via:

$$\begin{aligned} m &\sim N(\mu_m, \sigma_m^2), \\ b &\sim N(\mu_b, \sigma_b^2), \\ \sigma^2 &\sim \text{Exponential}(1/s_y), \end{aligned}$$

where m, b, σ^2 are independent.[5] The class discusses how to set the prior through its parameter values $\mu_m, \mu_b, \sigma_m^2, \sigma_b^2$ later in the activity. The instructor also discusses model assumptions for the Bayesian approach at a high level, stressing one of the fundamental differences between Frequentists and Bayesians: the intercept b and slope m parameters are treated as random variables with a distribution that represents our state of knowledge.

As with the Frequentist analysis, there are a number of questions asked of the class related to model summaries and general inference for the Bayesians. This time, however, students are shown the posterior distribution for the slope (m) and intercept (b) for a fitted model using Massachusetts data. The idea here is to get students thinking about how confident they should be in drawing conclusions from the model. After they have seen the distributions for these parameters, they are shown three different posterior distributions for the slope parameter so they can make sense of the results from the posterior:

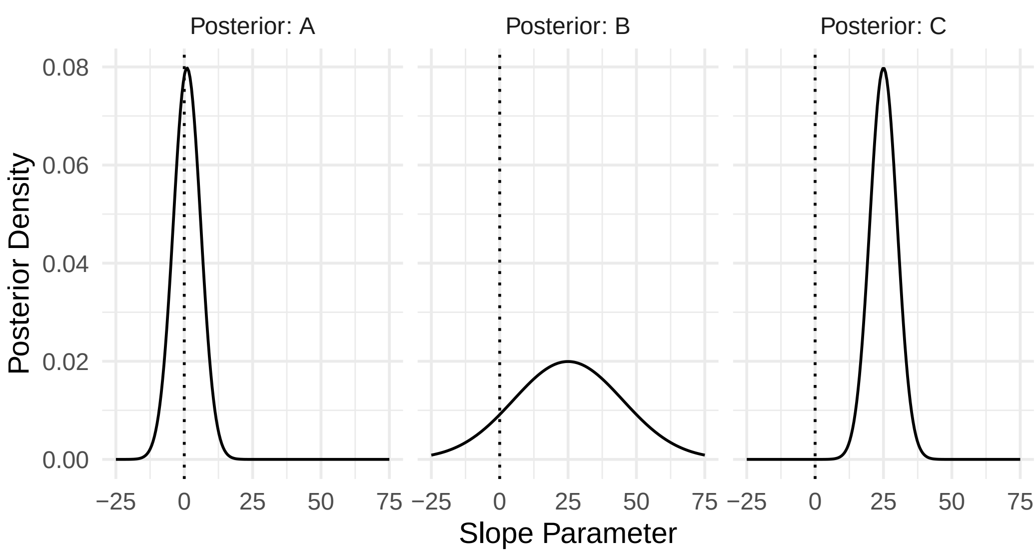


Figure 5: Figure 4: Notational Slope Posteriors for Bayesian Analysis

[5]Note that s_y is determined based on the standard deviation in the observed data. This is a way of *autoscaling* the prior.

Students are then asked a series of questions, a sample of which are below, to get them thinking about what the posterior distribution captures:

> Questions for the Class
>
> - Roughly, what fraction of *Posterior Distribution A* is greater than zero?
>   - Precisely 57.93%, so roughly 60%
> - Roughly, what fraction of *Posterior Distribution B* is greater than zero?
>   - Precisely 89.44%, so roughly 90%
> - Roughly, what fraction of *Posterior Distribution C* is greater than zero?
>   - Essentially 100%
> - Which posterior gives the *highest confidence* (*highest probability*) that the slope parameter is positive? How can you tell?
>   - Posterior C, as the probability is essentially 100% (as we identified above).

The students have the opportunity to look more closely at the slope parameter for Massachusetts and discuss it in the context of general inference. The students should recognize that a positive slope is in agreement with the hypothesis they are exploring before moving to the predictive portion of the Bayesian analysis, studying the posterior predictions.

To further stress the differences between the two approaches, the activity notes that a prior distribution must also be provided for all of the components of the model. These priors represent all of our prior knowledge about the real-world problem the students are trying to solve. It also ties all of the pieces together - a fitted Bayesian model is comprised of the *data + prior = posterior*. The activity walks the students through hypothetical scenarios where different prior distributions are used for the slope parameter when fitting the model with the Massachusetts data. This also shows the utility of a Bayesian approach when you have limited data, as it allows us to incorporate prior knowledge, a scenario we use in the activity (e.g., we assume the CEJS data is limited).

We employ a sequential Bayesian analysis by using the posterior from one analysis as the prior for a new analysis. We provide the following to the students in the Bayesian group and have the students pick a state's prior distribution for the rest of the activity:

> Pick a State
>
> Study the posteriors above carefully; you will use this as a *prior distribution* for the slope for the rest of the activity. This means you will combine *new data* with a *prior distribution* to form a new *posterior distribution* for the model parameters. The *prior distribution* should reflect your beliefs about what you think the slope parameter should be.
> Pick *one state for your group*, then come ask the instructor for your chosen state's

packet.

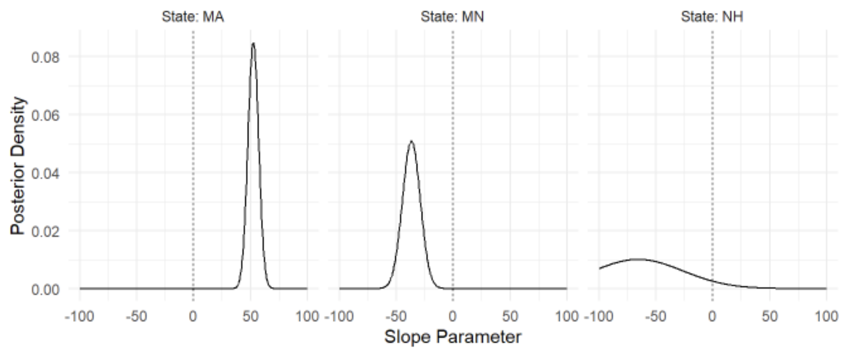


Figure 6: Example Posterior Choices from Bayesian Analysis

This represents a critical step in the activity for the Bayesians: they must pick their prior distribution based on their beliefs about what they think the slope parameter should be. This will subsequently be used with new data to form a posterior distribution later in the activity. Each of the results are placed in an envelope, only to be opened once selected by a group. The following figure represents an example of the sequential Bayesian analysis used in the full activity that students will have to choose from and interpret. On the left are the results showing the fitted lines and slope posteriors for Florida using priors from Massachusetts, Minnesota, and New Hampshire. On the right are the results showing the fitted lines and slope posteriors for Colorado, also using priors from Massachusetts, Minnesota, and New Hampshire.

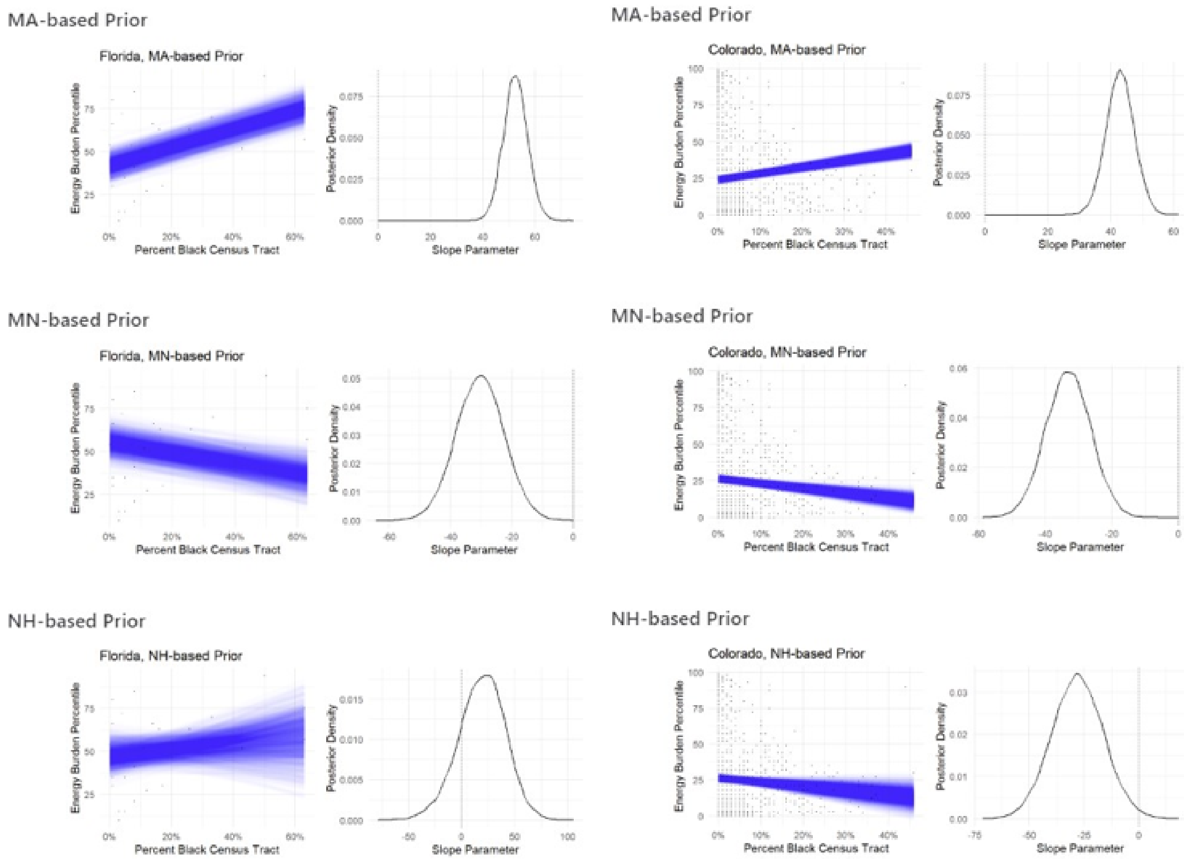


Figure 7: Comparison with Different Priors from Bayesian Analysis

Activity Closing

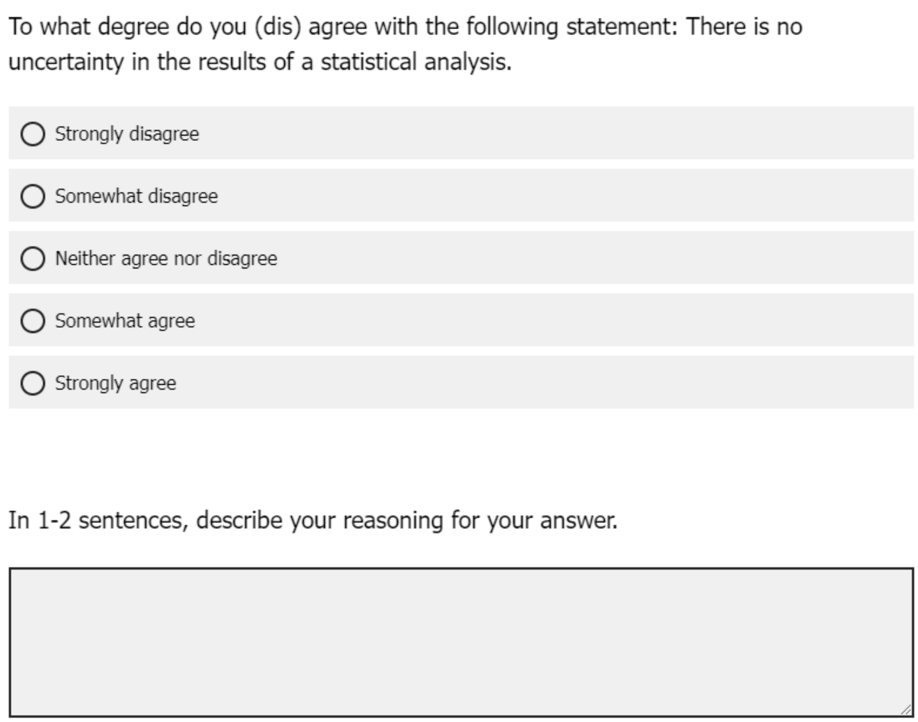
After students have had the opportunity to work through their respective applied analyses and discuss the questions in the associated document, the groups come back together for a full-class discussion. In addition to students jotting down any remaining questions they may have about each of the following questions, they are posed to the whole class for discussion and are tied to the learning objectives in Section :

- What can we say about our hypothesis?
- 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?

We use these questions for a few reasons. The first is so the entire class can hear the impressions of both groups regarding the statistical approach used in their analyses. The second is to play into the "controversy" or differences between the approaches to further engage students on the importance of assumptions for statistical conclusions. The instructor facilitates a debate between the two groups using the critical differences one-pager discussed in Section . Each group likely thinks their conclusions are "correct" based on their analyses, however it is important to point out that the controversy cannot be resolved; rather, the results from our analyses are conditional on the assumptions chosen, implying that choosing appropriate assumptions is critical to a sound analysis.

Evaluation

In addition to exposing students to different statistical paradigms, we are interested in students' awareness of Bayesian methods and their *epistemological framings* — their assumptions about the nature and accessibility of "truth" (Elby and Hammer 2010). Using pre- and post-pre-activity surveys, we measure students' self-reported familiarity of fundamental Bayesian ideas. Specifically, we measure their attitudes before the activity, as well as their change in attitudes after the activity, with respect to ideas around statistical inference. The pre-activity survey contains the consent form required by IRB, as well as a series of questions with Likert and open-ended responses that look like the following:


To what degree do you (dis) agree with the following statement: There is no uncertainty in the results of a statistical analysis.

- Strongly disagree
- Somewhat disagree
- Neither agree nor disagree
- Somewhat agree
- Strongly agree

In 1-2 sentences, describe your reasoning for your answer.

Figure 8: Example of Pre-Activity Survey

Once the activity is complete, we use a post-pre survey. The reason for using a post-pre design after the activity is straightforward. We want to capture changes in self-perceived attitudes about a topic by asking participants to consider where they think their beliefs were before the activity, followed by where they think they are now (see Hiebert and Magnusson (2014)).

The participants give themselves two ratings to capture this before and after reflection, as shown below.

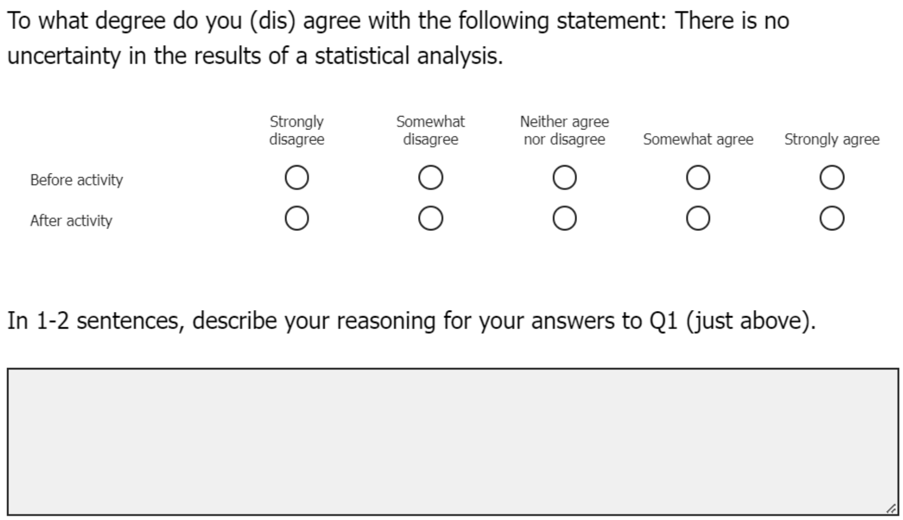


Figure 9: Example of Post-Pre-Activity Survey

We used a similar format for each of the learning objectives outlined in Section . The survey includes the following questions with Likert responses ranging from "Strongly disagree" to "Strongly agree":

1. *To what degree do you (dis) agree with the following statement: There is no uncertainty in the results of a statistical analysis.*
2. *To what degree do you (dis) agree with the following statement: The results of a statistical analysis should not depend on the analyst's assumptions.*
3. *To what degree do you (dis) agree with the following statement: I know how to relate statistical analysis to things in the real world.*

Each question that is associated with a learning objectives also gives students an open-ended opportunity to elaborate on their Likert responses. The survey ends by asking students the following open-ended question:

From the activity, what did you learn about the differences between Frequentist and Bayesian statistics? Please provide as much information as you can.

The post-pre survey is implemented after the activity closing, when students have had the opportunity to talk through the outcomes of each approach with their peers and the instructor. The results of the pilot implementation at the University of Denver and Olin College of Engineering are discussed next.

Preliminary Results

What follows are the preliminary results from the activity's implementation. At the University of Denver (responses on the left), there were n=12 pairs of valid pre- and post-pre responses. At Olin (responses on the right), there were n=25 pairs of valid responses across the surveys. The results for each question are shown below. For the DU responses on the left, the x-axis illustrates the Likert scale responses, while the y-axis represents the response for the pre-activity survey, the response for the pre-activity self-assessment in the post-activity survey, and the response for the post-activity self-assessment in the post-activity survey. The Olin responses on the right showcase the same information, albeit with the axes flipped and with the pre- and post-activity responses connected by lines. Each point on both figures represents one respondent.

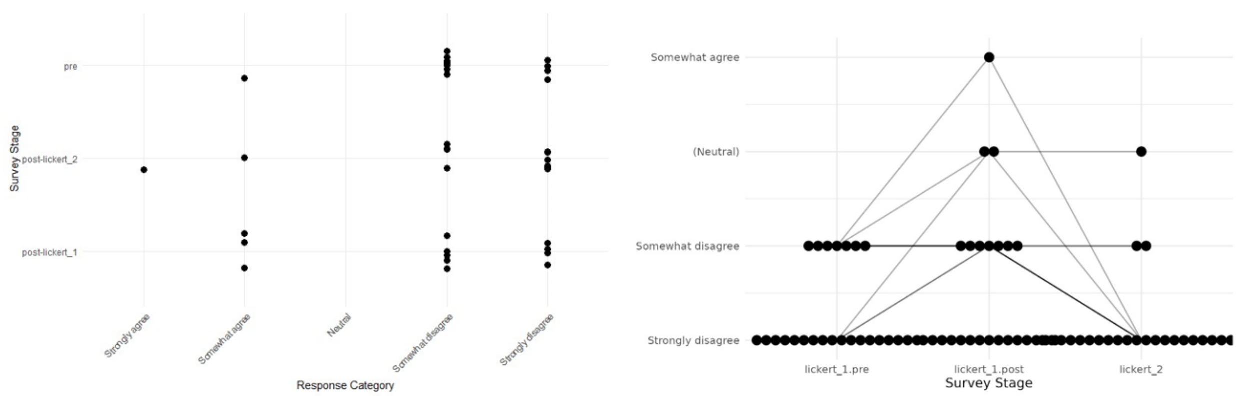


Figure 10: Question 1 - To what degree do you (dis) agree with the following statement: There is no uncertainty in the results of a statistical analysis.

For question one, students generally disagreed with the notion that there is not uncertainty associated with the results of statistical analysis across the different survey stages, though it is difficult to take away many meaningful insights from this particular question. Interestingly, a few respondents from DU trended towards the agree answers after the activity, suggesting they may have assumed Bayesians do not account for uncertainty in their results. Olin students' responses generally conform with what we would expect after the activity.

Figure 11: Question 2 - To what degree do you (dis) agree with the following statement: The results of a statistical analysis should not depend on the analyst's assumptions.

Question two had more variation with respect to the answers provided by both groups of students. Again, it is hard to discern many meaningful patterns around the pre- and post-activity battery of questions. We partially attribute this to poor question wording. We plan to change the question such that it reads "the results of a statistical analysis DO NOT depend on the analyst's assumptions" rather than what it is now.

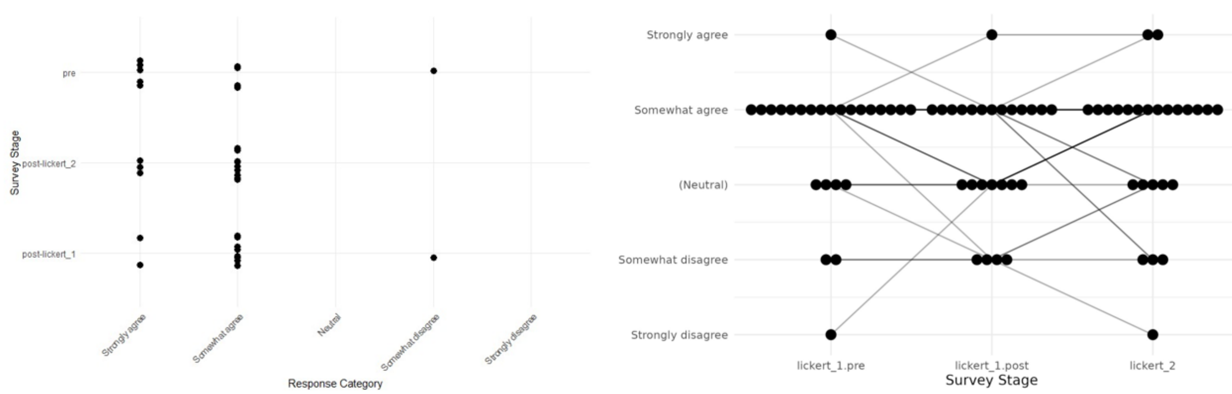


Figure 12: Question 3 - To what degree do you (dis) agree with the following statement: I know how to relate statistical analysis to things in the real world.

The third Likert question has students reflect on whether they know how to relate statistical analysis to things in the real world. For most students across the board, there was not a huge shift in attitudes. While there are a few students who appear to go between strongly and somewhat agree on this question, students perceive themselves as having a decent grasp of how to apply statistics to real-world problems.

Of interest are the responses to the open-ended question four in the post-activity survey. Recall the question asked, *"From the activity, what did you learn about the differences between Frequentist and Bayesian statistics? Please provide as much information as you can."* Responses ranged from simplistic, including "Frequentists aren't based on things or data from before while Bayesian is." to succinct, such as "Bayesian has a prior model, frequenting just takes the data as is." to relatively sophisticated: "The approaches are fundamentally different — Frequentists look at data alone while Bayesians consider prior information models in conjunction with their data. Both methods are valid modeling techniques, but Bayesian may be more useful to use when working with fewer data points (assuming a good prior is chosen) to help overcome deficiencies of minimal data." Finally, there were incorrect answers to this question. For example, one student stated that "Understanding how you use prior assumptions to shape new assumptions based on new data vs creating a new assumption for each new dataset."

It is important to note that these results should be interpreted with caution. While the survey questions are informative to a degree, the qualitative, open-ended responses initially provided far more insight into the learners' attitudes about inferential statistics and how their understanding changed over the course of the activity. Given the small sample size, we anticipate more robust findings after more iterations of the activity have been implemented.

Discussion

This activity bridges the gap between the common Frequentist approach oft taught in both undergraduate quantitative methods classes and the Bayesian paradigm, to which many students have not been exposed. We believe that non-statistics disciplines, such as public policy, are an arena with which Bayesian methods can be very beneficial. To bridge the gap, we use an applied approach, with an eye towards answering education research questions. While many teaching methods highlight the theoretical similarities and differences between Frequentists and Bayesians, our activity moves beyond by grounding the comparison in a real-data application, as well as measuring the impact of applying both frameworks in the classroom. In doing so, we hope to introduce students to a new way of using statistics, equipping them with the tool set and logical processes necessary to apply either the Frequentist or Bayesian (or both) approaches as they see fit.

Our activity uses an active learning approach, rather than passive lecture. Active learning has been shown to result in superior learning outcomes for students, particularly those from underrepresented groups (Freeman et al. 2014). We do this by structuring our activity around the 5E Model proposed by Duran and Duran (2004), which activates students' epistemological frames. The 5E Model is based on inquiry-based teaching, where students *engage, explore, explain, elaborate, and evaluate* statistical assumptions in an applied setting in order to promote broader impacts of inferential thinking with respect to statistics.

Our more speculative research goal—to promote more nuanced epistemological framings among students—has further potential impacts, and the preliminary results from our survey support this. Elby and Hammer (2010) argue that a "sophisticated" personal epistemology is actually achieved when one has access to multiple epistemological framings and can choose to switch between them based on what is productive for the context at hand. Students who can recognize and critique the assumptions underpinning their analyses (treating them as tentative), but carry out their analyses respecting those analyses (treating them as true) will likely be more effective as practicing statisticians. Getting students to recognize the importance of assumptions—and to practice adopting different assumptions—will be a critical first step in developing these multiple epistemological framings.

The activity described here is intended as a "minimum viable activity." Future avenues for extending this work include having students swap groups. Students in this activity only have the chance to engage deeply with only one of the two paradigms—Frequentist or Bayesian. A simple extension of the activity would be to have students re-do their analysis, but switch their approach. This will enable a more nuanced comparison between Frequentist and Bayesian approaches, which would add depth to the learning outcomes for students. The second extension is to create an interactive dashboard. Our initial design for the activity relies on students' ability to code in `R`. While this is feasible in our institutional contexts, it would limit the portability of the activity to contexts where programming skill is not so common. Thus, a no-code version of the activity would make it more accessible. Finally, another extension of the activity would be providing additional contexts and datasets. Our initial work uses a single context and dataset for the activity, however this could be redesigned to use a different context, which would promote the generalizability and impact of the activity.

References

Bates, Stephen R, and Laura Jenkins. 2007. "Teaching and Learning Ontology and Epistemology in Political Science." *Politics* 27 (1): 55–63.

Connelly, Steve, Dave Vanderhoven, Robert Rutherfoord, Liz Richardson, and Peter Matthews. 2021. "Translating Research for Policy: The Importance of Equivalence, Function, and Loyalty." *Humanities and Social Sciences Communications* 8 (1): 1–11.

Deke, John, Mariel Finucane, and Daniel Thal. 2022. "The BASIE (BAyeSian Interpretation of Estimates) Framework for Interpreting Findings from Impact Evaluations: A Practical Guide for Education Researchers. Toolkit. NCEE 2022-005." *National Center for Education Evaluation and Regional Assistance*.

Deslatte, Aaron, António Tavares, and Richard C Feiock. 2018. "Policy of Delay: Evidence from a Bayesian Analysis of Metropolitan Land-Use Choices." *Policy Studies Journal* 46 (3): 674–99.

Duran, Lena Ballone, and Emilio Duran. 2004. "The 5E Instructional Model: A Learning Cycle Approach for Inquiry-Based Science Teaching." *Science Education Review* 3 (2): 49–58.

Elby, Andrew, and David Hammer. 2010. "Epistemological Resources and Framing: A Cognitive Framework for Helping Teachers Interpret and Respond to Their Students' Epistemologies." *Personal Epistemology in the Classroom: Theory, Research, and Implications for Practice* 4 (1): 409–34.

Entwistle, Noel. 1997. "Contrasting Perspectives on Learning." *The Experience of Learning* 2: 3–22.

Fienberg, Stephen E. 2011. "Bayesian Models and Methods in Public Policy and Government Settings."

Freedman, David. 1997. "Some Issues in the Foundation of Statistics." *Topics in the Foundation of Statistics*, 19–39.

Freeman, Scott, Sarah L Eddy, Miles McDonough, Michelle K Smith, Nnadozie Okoroafor, Hannah Jordt, and Mary Pat Wenderoth. 2014. "Active Learning Increases Student Performance in Science, Engineering, and Mathematics." *Proceedings of the National Academy of Sciences* 111 (23): 8410–15.

Gill, Jeff, and Lee D Walker. 2005. "Elicited Priors for Bayesian Model Specifications in Political Science Research." *The Journal of Politics* 67 (3): 841–72.

Gill, Jeff, and Christopher Witko. 2013. "Bayesian Analytical Methods: A Methodological Prescription for Public Administration." *Journal of Public Administration Research and Theory* 23 (2): 457–94.

Gunn, Andrew. 2017. "Embedding Quantitative Methods by Stealth in Political Science: Developing a Pedagogy for Psephology." *Teaching Public Administration* 35 (3): 301–20.

Hiebert, Bryan, and Kris Magnusson. 2014. "The Power of Evidence: Demonstrating the Value of Career Development Services." *Career Development Practice in Canada: Perspectives, Principles, and Professionalism*, 489–530.

Jackman, Simon. 2000a. "Estimation and Inference Are Missing Data Problems: Unifying Social Science Statistics via Bayesian Simulation." *Political Analysis* 8 (4): 307–32.

———. 2000b. "Estimation and Inference via Bayesian Simulation: An Introduction to Markov Chain Monte Carlo." *American Journal of Political Science*, 375–404.

———. 2004. "Bayesian Analysis for Political Research." *Annual Review of Political Science* 7: 483–505.

———. 2009. *Bayesian Analysis for the Social Sciences*. John Wiley & Sons.

Kadane, Joseph B, and Anders Nordgaard. 2024. "Using Bayes Factors to Limit Forensic Testimony to Forensics: Composite Hypotheses." *Australian Journal of Forensic Sciences*, 1–17.

Luque, Carolina, and Juan Sosa. 2023. "Bayesian Analysis for Social Science Research." *arXiv Preprint arXiv:2306.11966*.

Samaniego, Francisco J. 2010. *A Comparison of the Bayesian and Frequentist Approaches to Estimation*. Vol. 24. Springer.

Shor, Boris, Joseph Bafumi, Luke Keele, and David Park. 2007. "A Bayesian Multilevel Modeling Approach to Time-Series Cross-Sectional Data." *Political Analysis* 15 (2): 165–81.

Uno, Gordon E. 1999. "Handbook on Teaching Undergraduate Science Courses: A Survival Training Manual." *(No Title)*.

Wagner, Kevin, and Jeff Gill. 2005. "Bayesian Inference in Public Administration Research: Substantive Differences from Somewhat Different Assumptions." *International Journal of Public Administration* 28 (1-2): 5–35.

Wibbens, Phebo D, Wesley Wu-Yi Koo, and Anita M McGahan. 2020. "Which COVID Policies Are Most Effective? A Bayesian Analysis of COVID-19 by Jurisdiction." *PloS*

One 15 (12): e0244177.