An Introduction to Bayesian Statistics for Psychological Research

Western Psychological Association 2024 Convention

April 28, 2024

Hyeri Hong, Ph.D.

Assistant Professor of Research and Statistics

California State University, Fresno

Email: hyerihong@mail.fresnostate.edu

Website: https://sites.google.com/mail.fresnostate.edu/prof-hyeri-hong?usp=sharing

ORCID:0000-0002-7576-2574

Researchgate: https://www.researchgate.net/profile/Hyeri-Hong

Alfonso J. Martinez

University of Iowa

Email: alfonso-martinez@uiowa.edu

Website: www.ajmquant.com

Twitter/X: AlfonsoMPsych

Part 2: Applications

Presented by Alfonso J. Martinez

Applications Outline

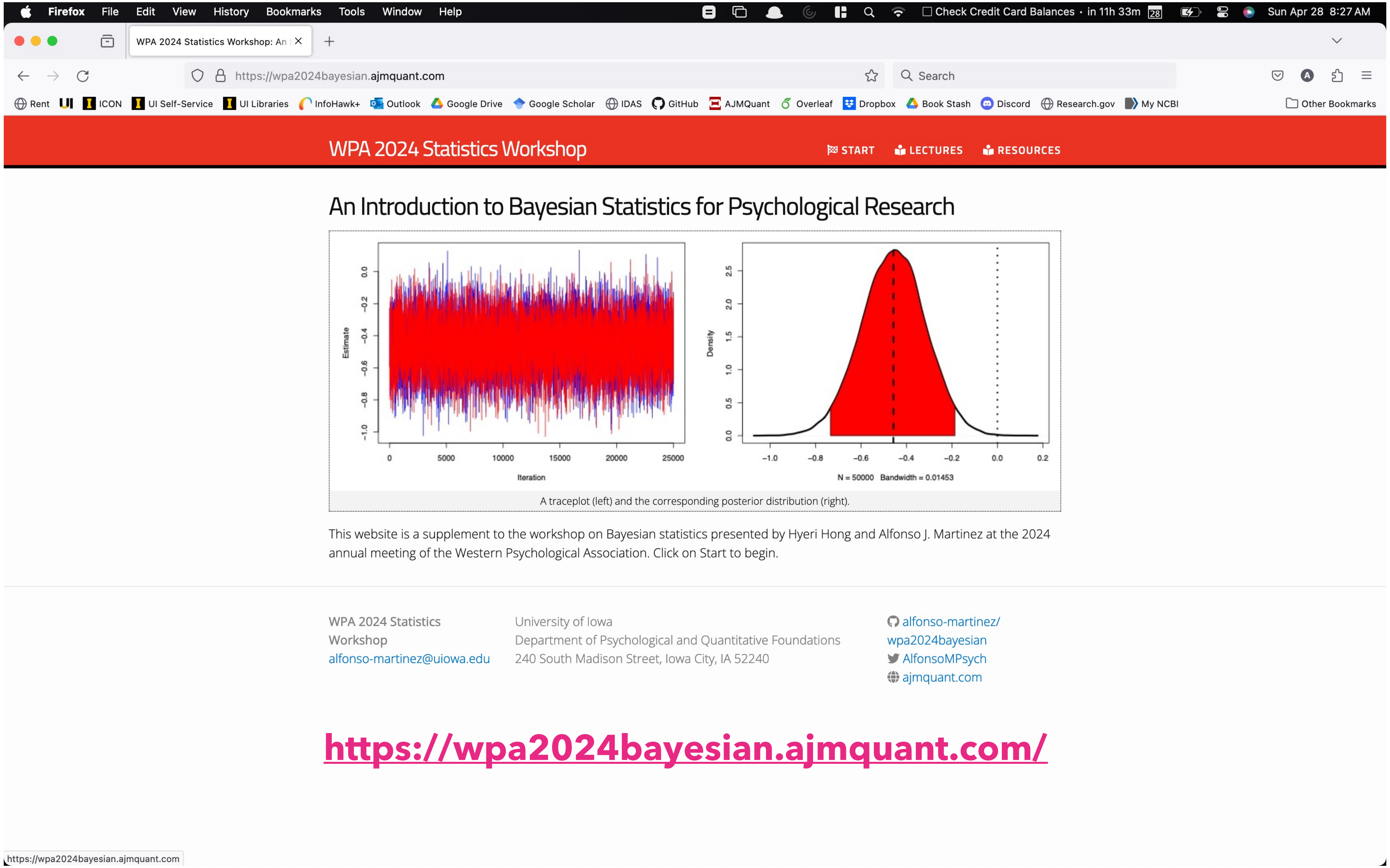
- **Example: linear regression**

- Three software programs:
  - SAS, Mplus, R

- Analysis presented loosely following the **WAMBS** checklist (Depaoli & Van de Schoot, 2017; *Psychological Methods*)

- We will focus on the basics, including model setup and interpreting the results
  - Topics we won't cover include model specification, parameterization, model identification, missing data, and model fit

Check Out Our Workshop Website!!



https://wpa2024bayesian.ajmquant.com/

- Contains slides and code for the example and different software programs

Software

R packages that can estimate Bayesian models

- Stan
- JAGS
- MCMCpack
- Nimble
- BayesianTools
- Blavaan (Bayesian SEM)
- **brms**
- A comprehensive list can be found here

ESTIMATOR = BAYES

PROC GENMOD

The Mplus and SAS files for the examples today will be uploaded to the workshop webpage in a few days

The WAMBS Checklist

Psychological Methods
2017, Vol. 22, No. 2, 240–261

1082-989X/17/$12.00 http://dx.doi.org/10.1037/met0000065

Improving Transparency and Replication in Bayesian Statistics: The WAMBS-Checklist

Sarah Depaoli
University of California, Merced

Rens van de Schoot
Utrecht University and North-West University

Bayesian statistical methods are slowly creeping into all fields of science and are becoming ever more popular in applied research. Although it is very attractive to use Bayesian statistics, our personal experience has led us to believe that naively applying Bayesian methods can be dangerous for at least 3 main reasons: the potential influence of priors, misinterpretation of Bayesian features and results, and improper reporting of Bayesian results. To deal with these 3 points of potential danger, we have developed a succinct checklist: the WAMBS-checklist (*W*hen to worry and how to *A*void the *M*isuse of *B*ayesian *S*tatistics). The purpose of the questionnaire is to describe 10 main points that should be thoroughly checked when applying Bayesian analysis. We provide an account of "when to worry" for each of these issues related to: (a) issues to check before estimating the model, (b) issues to check after estimating the model but before interpreting results, (c) understanding the influence of priors, and (d) actions to take after interpreting results. To accompany these key points of concern, we will present diagnostic tools that can be used in conjunction with the development and assessment of a Bayesian model. We also include examples of how to interpret results when "problems" in estimation arise, as well as syntax and instructions for implementation. Our aim is to stress the importance of openness and transparency of all aspects of Bayesian estimation, and it is our hope that the WAMBS questionnaire can aid in this process.

Keywords: Bayesian estimation, prior, sensitivity analysis, convergence, Bayesian checklist

Supplemental materials: http://dx.doi.org/10.1037/met0000065.supp

Bayesian statistical methods are slowly creeping into all fields of science and are becoming ever more popular in applied research. Figure 1 displays results from a literature search in Scopus using the term "Bayesian estimation" and, as can be seen, the number of empirical peer-reviewed articles using Bayesian estimation is on the rise. This increase is likely due to recent computational advancements and the availability of Bayesian estimation methods in popular software and programming languages like WinBUGS and OpenBUGS (Lunn, Thomas, Best, & Spiegelhalter, 2000), MlWiN (Browne, 2009), AMOS (Arbuckle, 2006), M*plus* (Muthén & Muthén, 1998–2015), BIEMS (Mulder, Hoijtink, & de Leeuw, 2012), JASP (Love et al., 2015) via the BayesFactor package in R, which is also a standalone Bayesian package (Morey, Rouder, & Jamil, 2015), SAS (SAS Institute Inc., 2002–2013), and STATA (StataCorp., 2013). Further, there are various packages in the R programming environment (Albert, 2009) such as STAN (Stan Development Team, 2014) and JAGS (Plummer, 2003) that implement Bayesian methods.

When to Use Bayesian Statistics

There are (at least) four main reasons why one might choose to use Bayesian statistics. First, some complex models simply cannot be estimated using conventional statistics (see, e.g., Muthén & Asparouhov, 2012; Kruschke, 2010, 2011; Wetzels, Matzke, Lee, Rouder, Iverson & Wagenmakers, 2011). Further, some models (e.g., mixture or multilevel models) require Bayesian methods to improve convergence issues (Depaoli & Clifton, 2015; Skrondal & Rabe-Hesketh, 2012), aid in model identification (Kim, Suh, Kim, Albanese, & Langer, 2013), and produce more accurate parameter estimates (Depaoli, 2013, 2014). Second, many scholars prefer Bayesian statistics because they believe population parameters should be viewed as random (see, e.g., Dienes, 2011; van de Schoot et al., 2011). Third, with Bayesian statistics one can incorporate (un)certainty about a parameter and update this knowledge through the prior distribution. Fourth, Bayesian statistics is not based on large samples (i.e., the central limit theorem) and hence may produce reasonable results even with small to moderate sample sizes, especially when strong and defensible prior knowledge is available (Hox, van de Schoot, & Matthijsse, 2012; Moore et al., 2015; van de Schoot, Broere, Perryck, Zondervan-Zwijnenburg, & van Loey, 2015; Zhang, Hamagami, Wang, Grimm, & Nesselroade, 2007).

This article was published Online First December 21, 2015.

Sarah Depaoli, Department of Psychological Sciences, University of California, Merced; Rens van de Schoot, Department of Methods and Statistics, Utrecht University, and Optentia Research Program, Faculty of Humanities, North-West University.

The second author was supported by a grant from the Netherlands organization for scientific research: NWO-VENI-451-11-008.

Correspondence concerning this article should be addressed to Sarah Depaoli, Psychological Sciences, University of California, Merced, 5200 N. Lake Road, Merced, CA, 95343. E-mail: sdepaoli@ucmerced.edu

- ❑ **Questionnaire designed to guide researchers through a Bayesian analysis**
- ❑ 10 step checklist
- ❑ Four categories
  - ❑ Considerations before model estimation
  - ❑ Considerations after model estimation but before inspection of results
  - ❑ Understanding the influence of priors
  - ❑ Interpretation of results

Depaoli, S., & Van de Schoot, R. (2017). Improving Transparency and Replication in Bayesian Statistics: The WAMBS-Checklist. *Psychological Methods, 22*(2), 240-261. http://dx.doi.org/10.1037/met0000065

The WAMBS Checklist

THE WAMBS-CHECKLIST
When to worry, and how to Avoid the Misuse of Bayesian Statistics
DEPAOLI & VAN DE SCHOOT (2016)

	Did you show your supervisor…?	Should you worry?	Should you consult an expert?
TO BE CHECKED BEFORE ESTIMATING THE MODEL			
Point 1: Do you understand the priors?	Table 1	YES / NO	YES / NO
TO BE CHECKED AFTER ESTIMATION BUT BEFORE INSPECTING MODEL RESULTS			
Point 2: Does the trace-plot exhibit convergence?	Table 2, column 2	YES / NO	YES / NO
Point 3: Does convergence remain after doubling the number of iterations?	Table 4, columns 2, 3 (i) and akin to Table 3	YES / NO	YES / NO
Point 4: Does the histogram have enough information?	Table 2, column 3	YES / NO	n/a
Point 5: Do the chains exhibit a strong degree of autocorrelation?	Table 2, column 4	YES / NO	YES / NO
Point 6: Does the posterior distribution make substantive sense?	Table 2, column 5	YES / NO	YES / NO
UNDERSTANDING THE EXACT INFLUENCE OF THE PRIORS			
Point 7: Do different specifications of the multivariate variance priors influence the results?	Table 4, columns 2, 3 (ii)	YES / NO	YES / NO
Point 8: Is there a notable effect of the prior when compared with non-informative priors?	Table 4, columns 2, 3 (iii)	NEVER	n/a
Point 9: Are the results stable from a sensitivity analysis?	Sensitivity analysis akin to Table 5 or Figure 4	NEVER	YES / NO
AFTER INTERPRETATION OF MODEL RESULTS			
Point 10: Is the Bayesian way of interpreting and reporting model results used? *(a) Also report on: missing data, model fit and comparison, non-response, generalizability, ability to replicate, etc.*	Text – see Appendix	YES / NO	YES / NO

Figure taken from Depaoli, S., & Van de Schoot, R. (2017). Improving Transparency and Replication in Bayesian Statistics: The WAMBS-Checklist. *Psychological Methods, 22*(2), 240-261. http://dx.doi.org/10.1037/met0000065

Example 1: Linear Regression

Linear Regression

- The goal of a **linear regression analysis** is to describe the influence a set of independent variables (predictors) $X_1, \dots, X_p$ have on a continuous dependent (response) variable Y_i

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots \beta_p X_{ip} + \epsilon_i$$

- The response variable Y_i is modeled as a linear combination of the predictors $X_1, \dots, X_p$

- Two types of parameters in a linear regression model
  - The $\boldsymbol{\beta}$ terms are the **regression coefficients** that describe the influence a given predictor has on the outcome
  - A random error term $\epsilon_i \sim N(0, \sigma^2)$ that captures any random source of variability → main interest is in σ^2 (**amount of variability** across a sample of n observations)

- In real data applications, $\boldsymbol{\beta}$ and σ^2 are **unknown** and must be estimated from the data available

!! Even though β and σ^2 are unknown they are not random !!

Aside: Non-Bayesian Estimation of the Linear Regression Model

- Assuming normality of the residuals, i.e., $\epsilon_i \sim N(0, \sigma^2)$ then Y_i will be **normally distributed** random variable with mean μ_i and variance σ^2

$$Y_i \mid \boldsymbol{X}_i \sim N(\mu_i, \sigma^2)$$

where $\mu_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots \beta_p X_{ip}$

- In least squares (LS) estimation, the beta coefficients $\boldsymbol{\beta}$ are estimated by **minimizing** the residual sums of squares with respect to $\boldsymbol{\beta}$:

$$S(\boldsymbol{\beta}) = \sum_{i=1}^{n} (y_i - \mu_i)^2$$

Note: minimizing $S(\boldsymbol{\beta})$ is equivalent to maximizing the likelihood function

- Under LS, we can obtain an explicit formula for the regression coefficients and residual variance:

$$\widehat{\boldsymbol{\beta}}_{LS} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{y}$$

$$\hat{\sigma}_{LS}^2 = \frac{S(\boldsymbol{\beta})}{n - p - 1}$$

Motivating Bayes: A Simulation Study

- Example of LS using simulated data. The model is

$$Y_i = \beta_0 + \beta_1 X_{i1} + \epsilon_i$$

- Data generating specifics → $n = 150;\ \beta_0 = 0;\ \beta_1 = 1;\ \sigma^2 = 1$

Results from Least Squares Estimation (via `lm`)

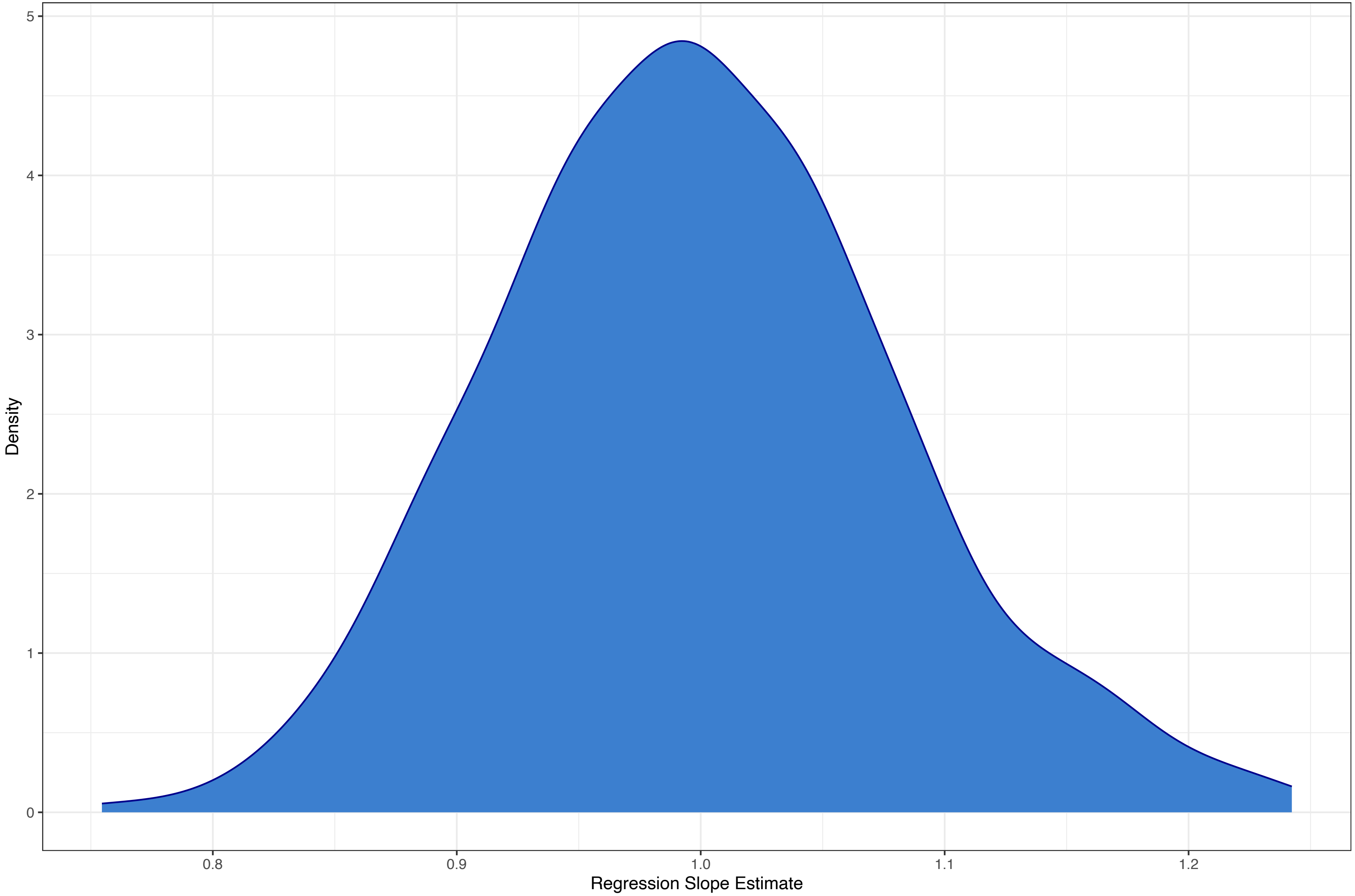
	Estimate	SE	95% Conf. Int.
β_0	0.123	0.082	(-0.039, 0.285)
β_1	0.917	0.073	(0.774, 1.061)
σ^2	1.006		

These β estimates are the values that **minimize** the residual sums of squares $S(\boldsymbol{\beta})$

The standard errors and confidence intervals are interpreted with respect to the **sampling distributions** of the parameters

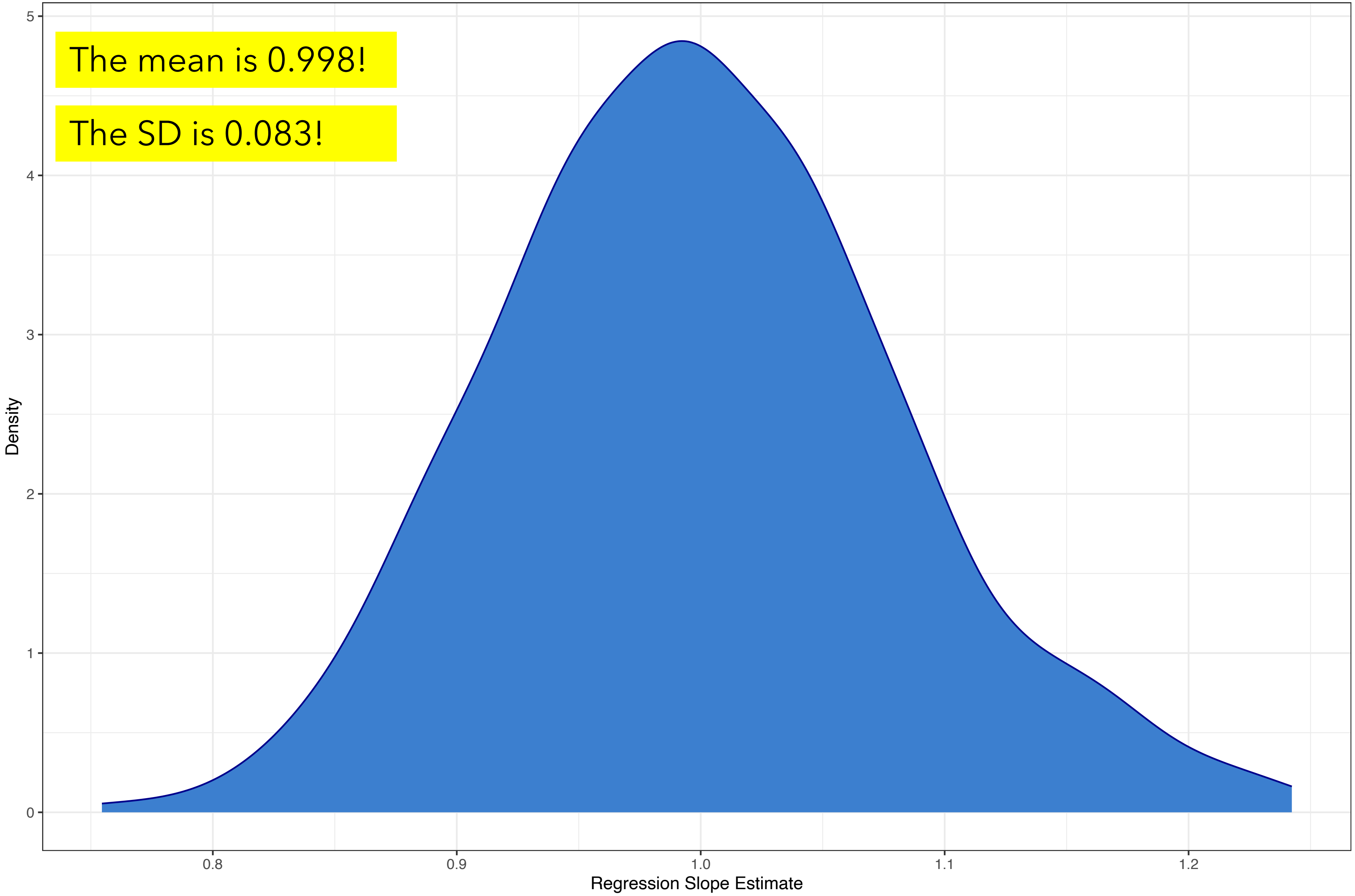
- Conceptually, it can be helpful to think of repeated sampling as **repeating an experiment** many, many times under the exact same conditions but with a new sample each time (and parameter estimates are saved each time as well) → **histogram is an estimate of the theoretical sampling distribution**
- Under **repeated sampling**, the SE is the standard deviation of the sampling distribution
- Under **repeated sampling**, 95% of CIs would contain the true population values

Motivating Bayes: The Sampling Distribution of β_1



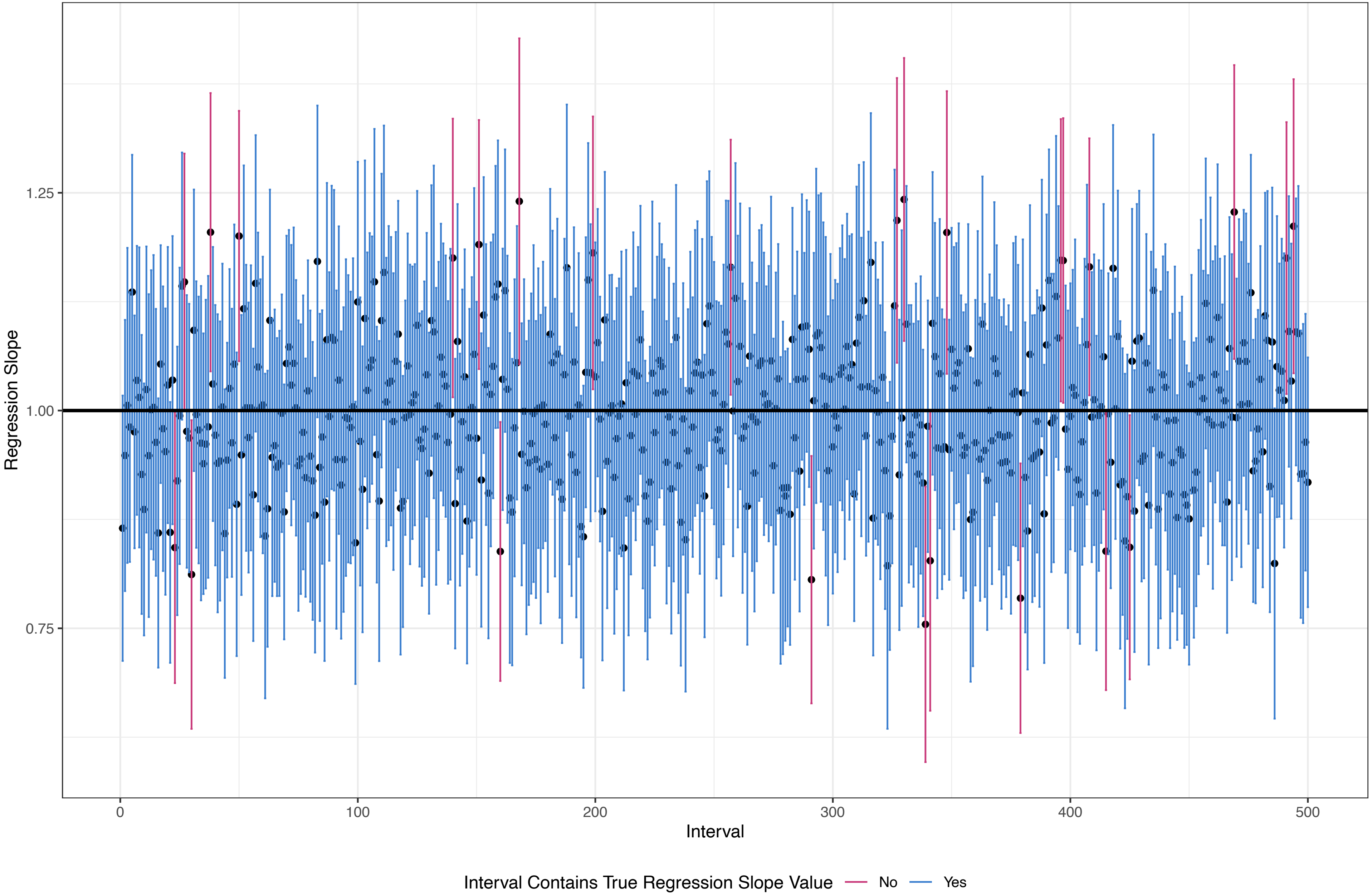
- Here the **"experiment" was repeated 500 times**, each time with a new dataset (but same underlying model)
- I saved the estimates of β_1 after each replication until all 500 replications were complete
- This is the histogram of the 500 β_1 estimates

Motivating Bayes: The Sampling Distribution of β_1



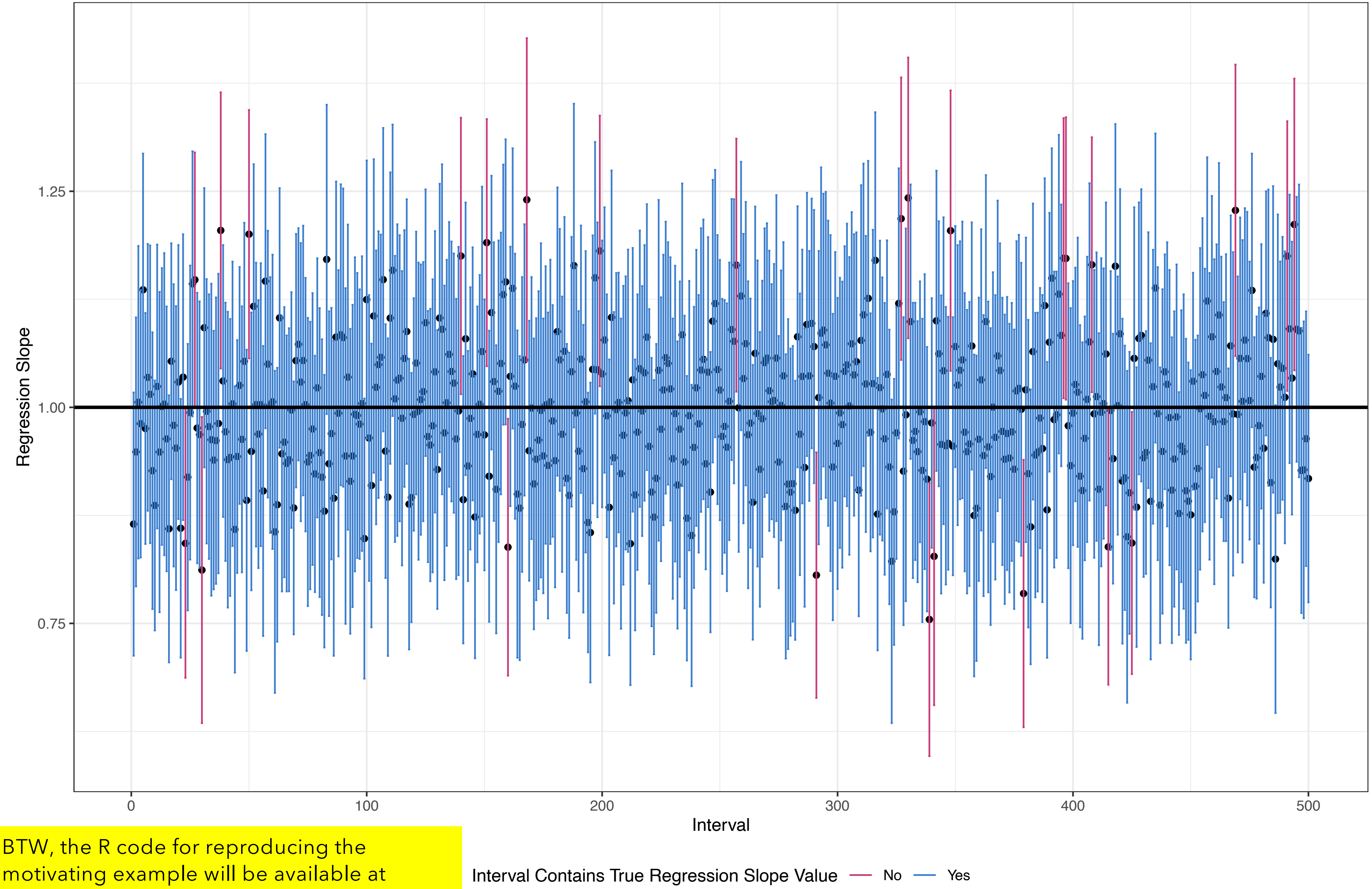
- The mean is **very close** to the true value $\beta_1 = 1$
- The SD is **somewhat close** to the SE value from previous slide (it gets better as n gets larger)
- BTW, the theoretical SD is $\frac{1}{\sqrt{n}} \rightarrow \frac{1}{\sqrt{150}} = 0.08164966$
- Thus, the **mean and SE from the previous slide are estimates of the mean and SD from this hypothetical sampling distribution!**

Motivating Bayes: What About the Confidence Interval?



- Here the **"experiment" was repeated 500 times**, each time with a new dataset (but same underlying model)
- Each dot is the estimate of β_1 and the bars represent the 95% CIs intervals *based on that* replication
- Guess how many of the 500 datasets had intervals that contained $\beta_1 = 1$?

Motivating Bayes: What About the Confidence Interval?



- Here the **"experiment" was repeated 500 times**, each time with a new dataset (but same underlying model)
- Each dot is the estimate of β_1 and the bars represent the 95% CIs intervals *based on that* replication
- Guess how many of the 500 datasets had intervals that contained $\beta_1 = 1$?

474 out of 500 (94.8%)!

BTW, the R code for reproducing the motivating example will be available at **https://wpa2024bayesian.ajmquant.com/**

Motivating Bayes: Recap

- As we just saw, the point estimates, SE, and CIs are interpreted with respect to a **hypothetical sampling distribution** which relies on the notion of **repeated sampling**

- The idea of an **infinitely repeating "experiment"** is not intuitive in many contexts
  - Intuitive if the "experiment" is flipping a coin
  - Not intuitive if the "experiment" is a study that investigates effects of environmental factors on mental health

- Also, notice that **at no point** did I mention that a researcher has the ability to **incorporate their domain knowledge and expertise** into the analysis

- Bayesian estimation [of the linear regression model] addresses these issues

Bayesian Estimation of the Linear Regression Model

- Recall: the parameters of interest in the linear regression model are $\boldsymbol{\beta}$ and σ^2

- In Bayesian estimation, we treat $\boldsymbol{\beta}$ and σ^2 as **random variables that have distributions**

The goal is to update our beliefs about $\boldsymbol{\beta}$ and σ^2 **in light of the data we collected**

- This is encoded in **Bayes' theorem**

$$P(\boldsymbol{\beta}, \sigma^2 \mid \boldsymbol{y}) \propto L(\boldsymbol{\beta}, \sigma^2; \boldsymbol{y}) P(\boldsymbol{\beta}) P(\sigma^2)$$

Posterior distribution distribution of $\boldsymbol{\beta}$ and σ^2 given data $\boldsymbol{y}$

Likelihood function (the information contained in the data)

Prior distributions of $\boldsymbol{\beta}$ and σ^2, respectively

Everything we want to know about $\boldsymbol{\beta}$ and σ^2 based on the available data (and our prior beliefs of $\boldsymbol{\beta}$ and σ^2) is contained in the posterior

Steps of a Bayesian Analysis

- ❑ Specify the likelihood
  - ❑ Linear regression model
- ❑ Identify the parameters of interest
  - ❑ $\boldsymbol{\beta}$ and σ^2
- ❑ Specify the priors
- ❑ Specify and estimate the model with statistical software
- ❑ Check diagnostics to see if results are trustworthy/reasonable
- ❑ Interpret results (create tables, graphs, construct credible intervals, etc.)

Aside: For the linear regression model, there are prior distributions that will give us closed form (aka "nice") posteriors of $\boldsymbol{\beta}$ and σ^2 [conjugate priors] but this topic is beyond the scope of this presentation (see Gelman et al., 2004 for more details)

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). Bayesian data analysis (2nd ed.). London, UK: Chapman & Hall.

Application to Counseling Psychology: Healthcare Career Interests

- Yanez, G. F., **Martinez, A. J.**, Ali, S. R., & Son, Y. (**under review**). Sociopolitical development and healthcare interests among rural youth: Is gender a moderator?
  - Note: As the paper is currently undergoing peer-review, the data used in this application is a **simulated** version of the real dataset.

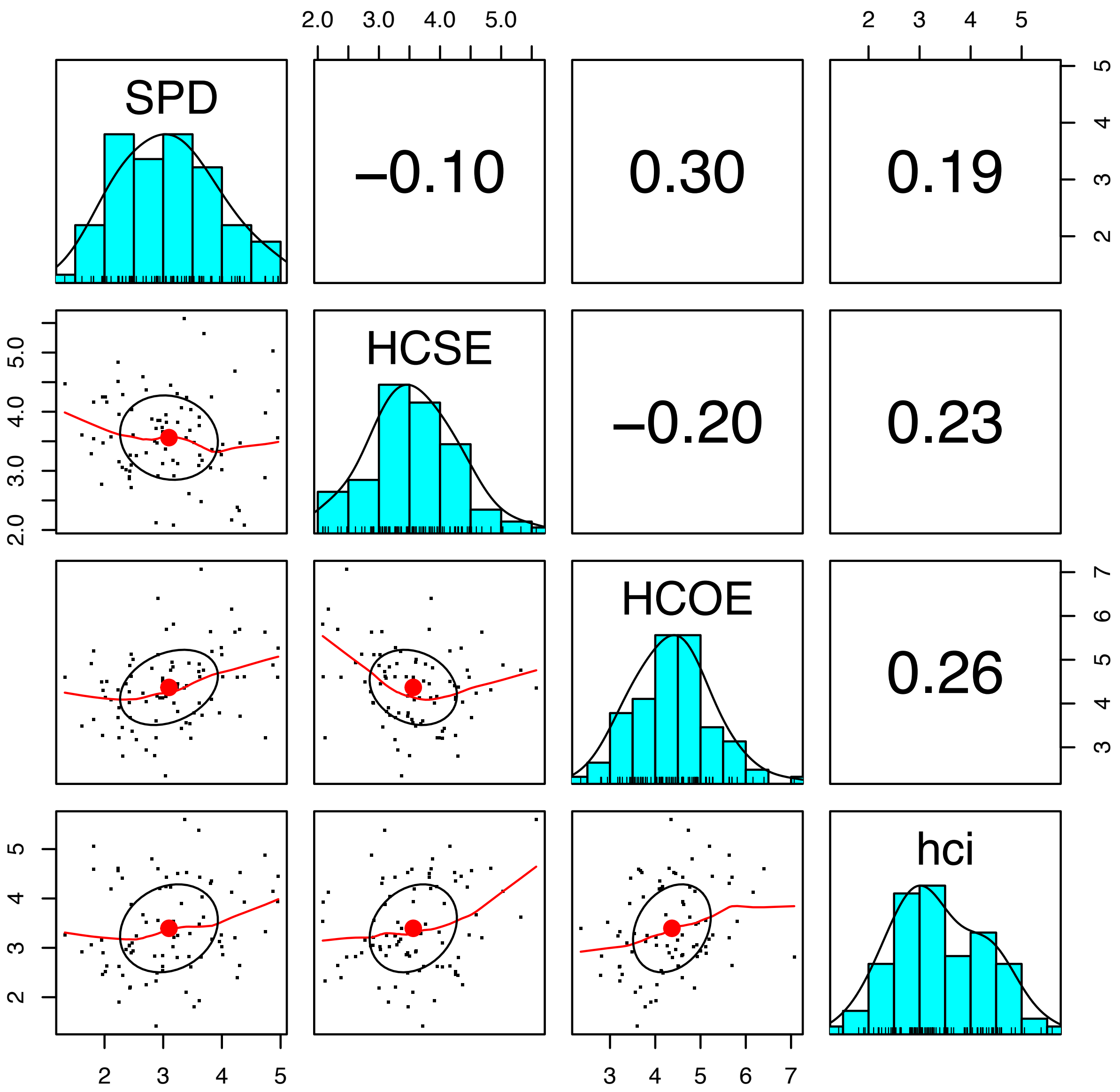
- Replication of Ali et al. (2021) which examined differences in sociopolitical development and healthcare career-related outcomes in rural youth

- $n = 85$ **8th graders** from a middle school in the rural Midwest participated in the study

- This example is a simplified version of the models tested in the paper
  - Four variables: sociopolitical development (SPD), healthcare career interest (HCI), healthcare outcome expectations (HCOE), healthcare self-efficacy (HCSE)

Ali, S. R., Loh Garrison, Y., Cervantes, Z. M., & Dawson, D. A. (2021). Sociopolitical development and healthcare career interest, self-efficacy, and outcome expectations among rural youth. *The Counseling Psychologist*, *49*(5), 701-727.

The Data



Application to Counseling Psychology: Healthcare Career Interests

- **The linear regression model for this analysis is**

$$\text{HCI}_i = \beta_0 + \beta_1\text{SPD}_i + \beta_2\text{HCSE}_i + \beta_3\text{HCOE}_i + \epsilon_i$$

Symbol	Variable Name	Variable Description	Parameter
HCI	Healthcare career interest	Degree to which students are interested in pursuing healthcare careers	
SPD	Sociopolitical development	Youth's self-perception of their ability to make an impact on policy decisions at the community level	β_1
HCSE	Self-efficacy	Confidence in doing healthcare career-related tasks	β_2
HCOE	Outcome expectations	Students' beliefs about how their actions will impact their future school and career choices	β_3

Note: all variables come from self-report measures; all variables except SPD were measured on a 6-point Likert scale (SPD was measured on a 5-point Likert scale)

Ali, S. R., Loh Garrison, Y., Cervantes, Z. M., & Dawson, D. A. (2021). Sociopolitical development and healthcare career interest, self-efficacy, and outcome expectations among rural youth. *The Counseling Psychologist*, *49*(5), 701-727.

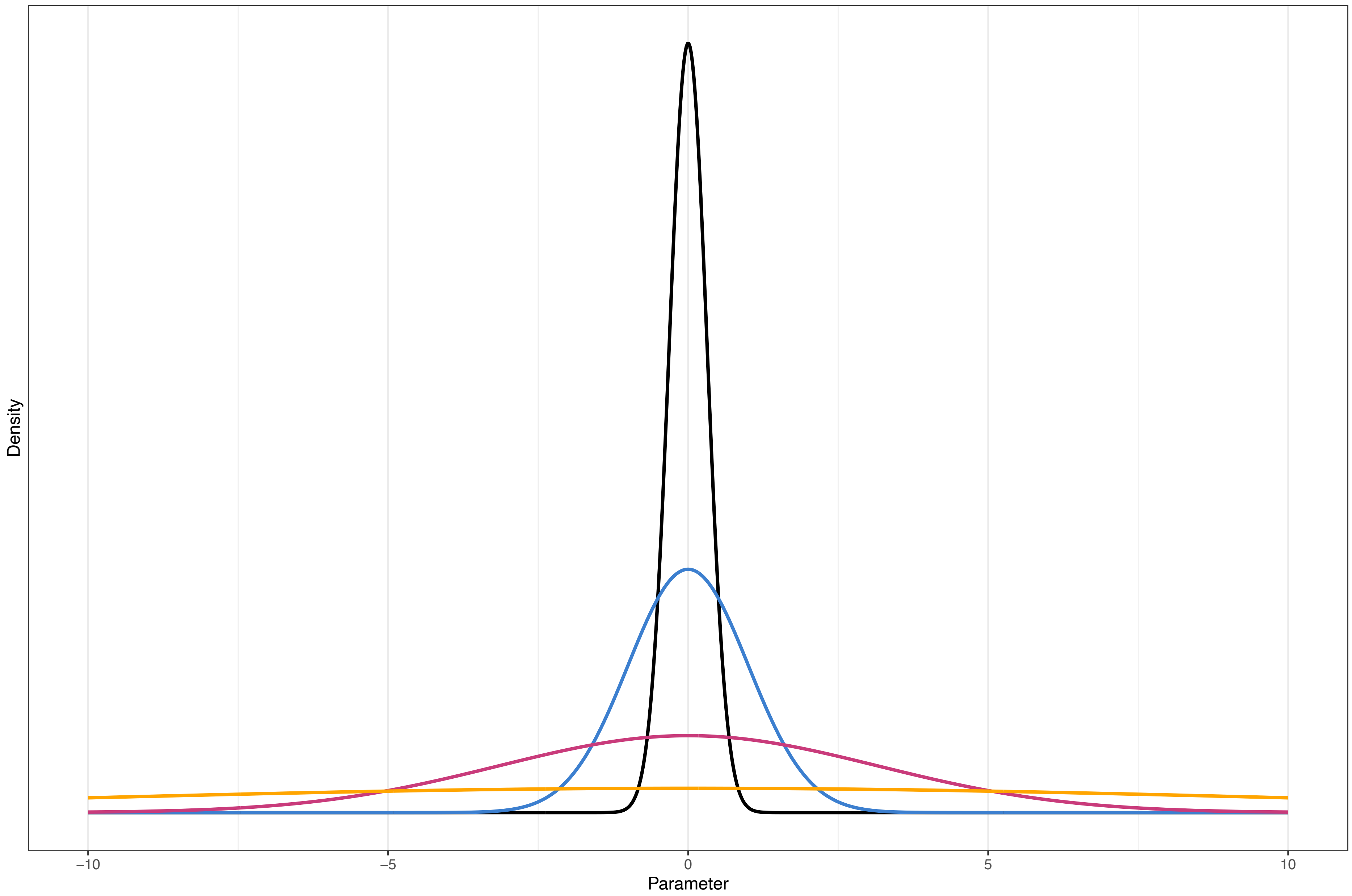
Choosing priors

- There are five parameters for which we need to specify priors for:
  - β_0 (intercept)
  - β_1 (effect of SPD on HCI)
  - β_2 (effect of HCSE on HCI)
  - β_3 (effect of HCOE on HCI)
  - σ^2 (residual variance; variance unexplained by the other three predictors)

- In general, priors should match the support of the parameters
  - For the $\boldsymbol{\beta}$ parameters, any distribution with support over the real line may be reasonable (e.g., a normal distribution)

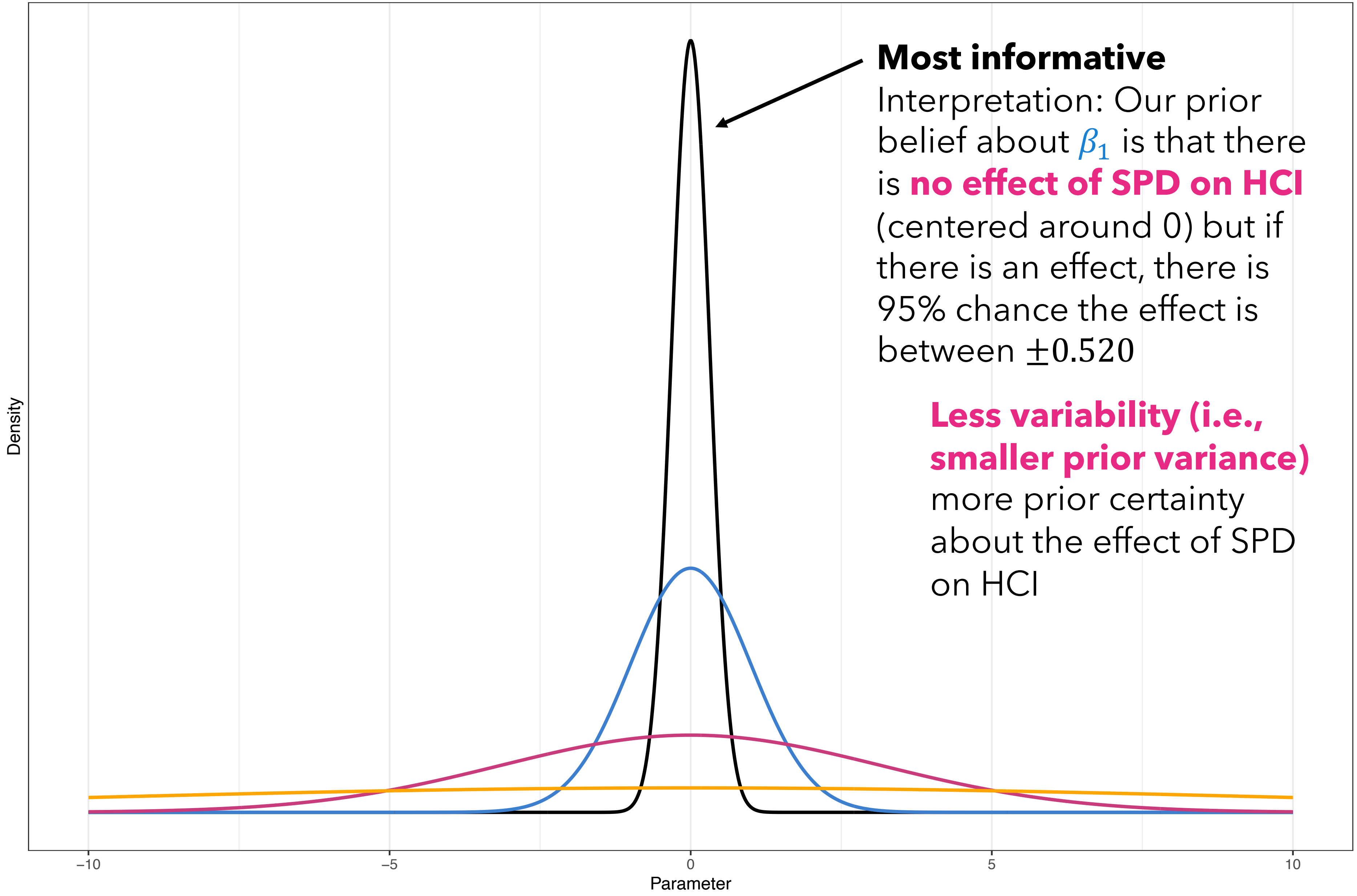
  - Since σ^2 shouldn't be negative, our prior should have support over the positive real line (e.g., a normal distribution is not appropriate for σ^2)

Examples of Different Normal Priors for β_1 (Effect of SPD on HCI)



Prior	95%
$N(0, 0.1)$	± 0.520
$N(0, 1)$	± 1.644
$N(0, 0.10)$	± 5.201
$N(0, 100)$	± 16.448

Examples of Different Normal Priors for β_1 (Effect of SPD on HCI)

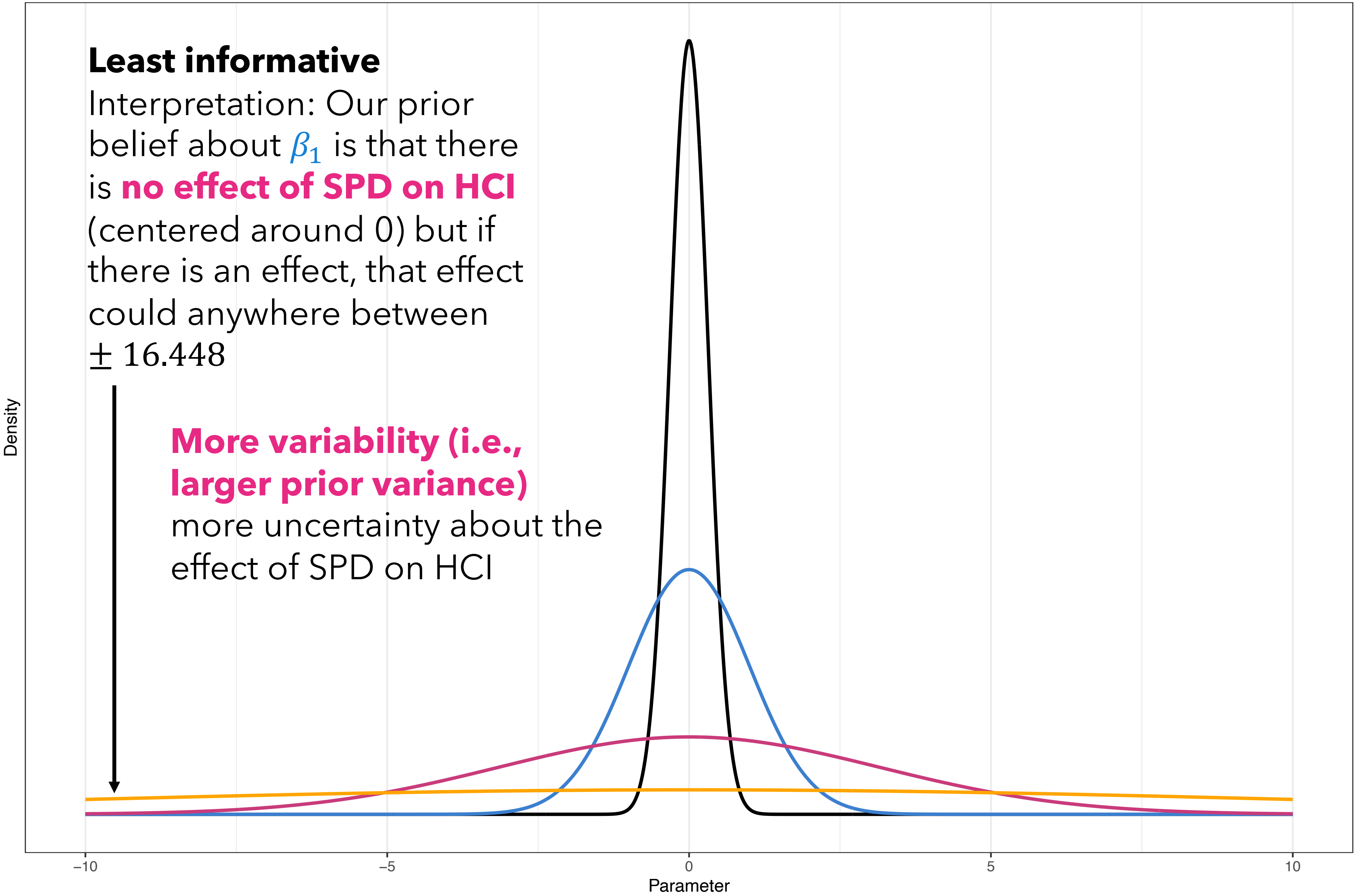


Most informative
Interpretation: Our prior belief about β_1 is that there is **no effect of SPD on HCI** (centered around 0) but if there is an effect, there is 95% chance the effect is between ± 0.520

Less variability (i.e., smaller prior variance) more prior certainty about the effect of SPD on HCI

Prior	95%
$N(0, 0.1)$	± 0.520
$N(0, 1)$	± 1.644
$N(0, 0.10)$	± 5.201
$N(0, 100)$	± 16.448

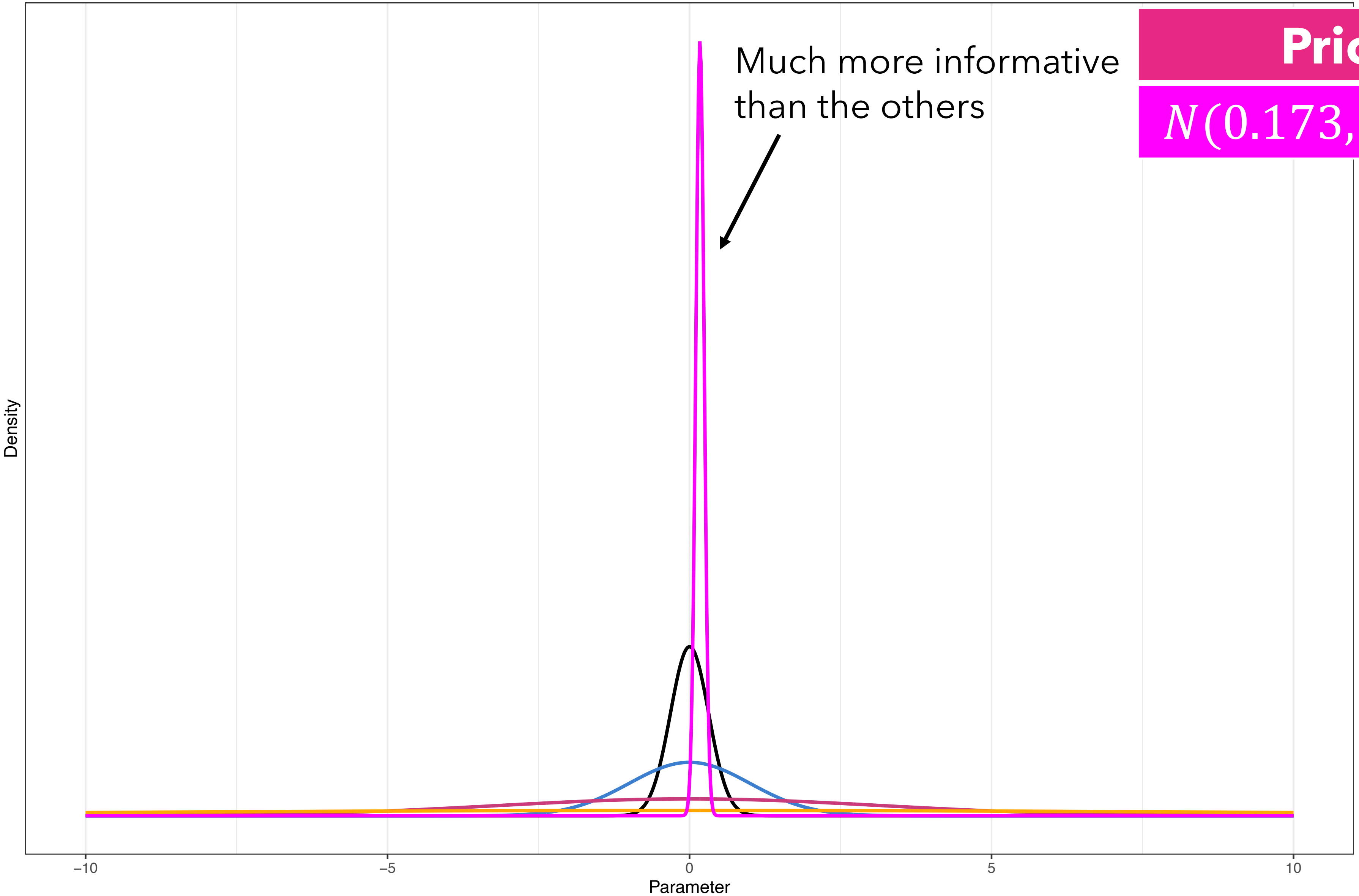
Examples of Different Normal Priors for β_1 (Effect of SPD on HCI)



Prior	95%
$N(0, 0.1)$	± 0.520
$N(0, 1)$	± 1.644
$N(0, 0.10)$	± 5.201
$N(0, 100)$	± 16.448

BTW, the values we specify for the priors are called hyperparameters – you can put priors on these too!

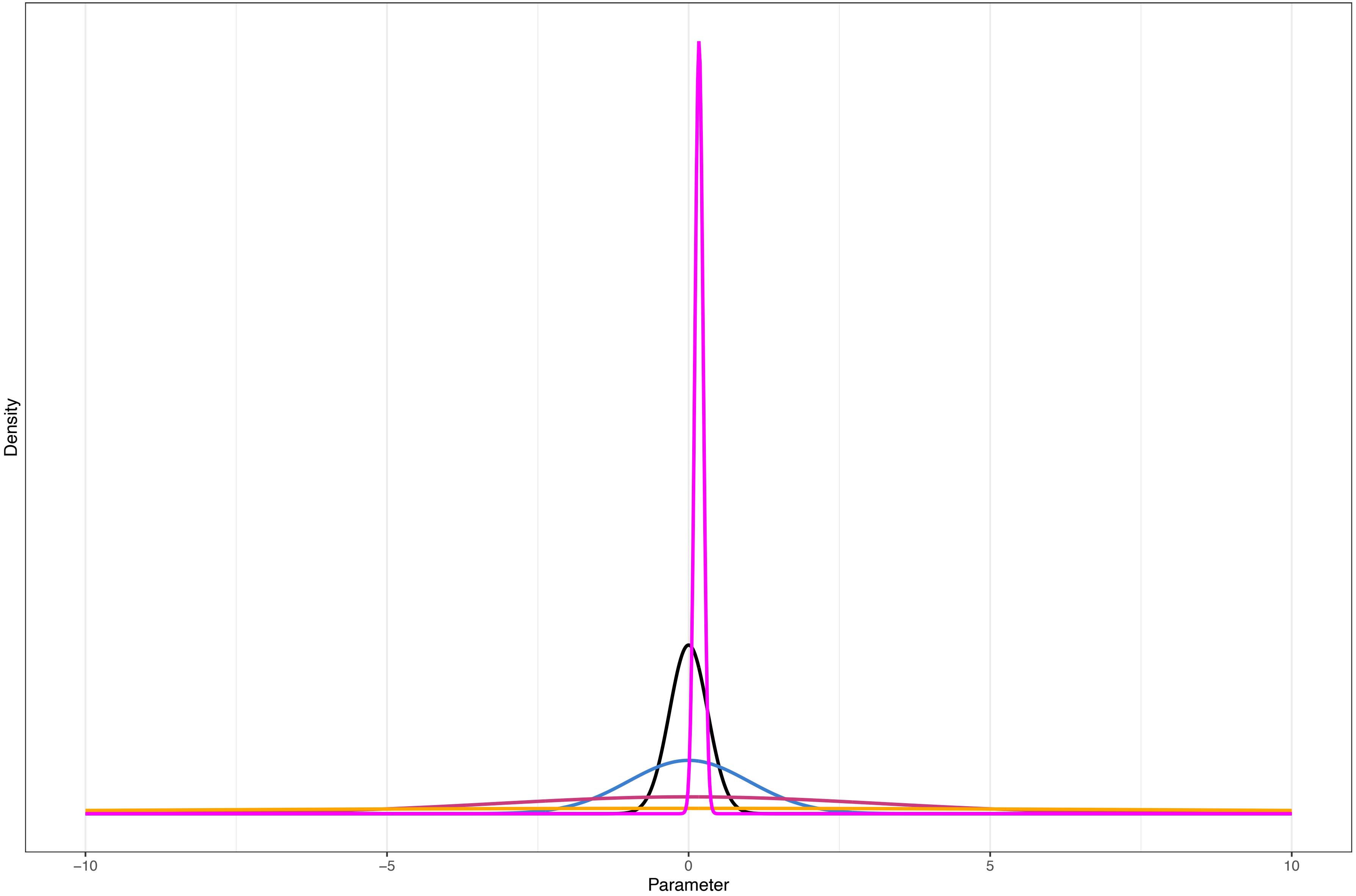
Examples of Different Normal Priors for β_1 (Effect of SPD on HCI)



Prior	95%
$N(0.173, 0.69^2)$	$(0.059, 0.286)$

- This prior was constructed by using the results of previous research [e.g., Ali et al. (2021)]
- Specifically, we took the point estimate of 0.173 as the prior mean and the bootstrap SE of 0.69 to define the prior variance of 0.69^2
- **Meta-analysis approach**

Examples of Different Normal Priors for β_1 (Effect of SPD on HCI)

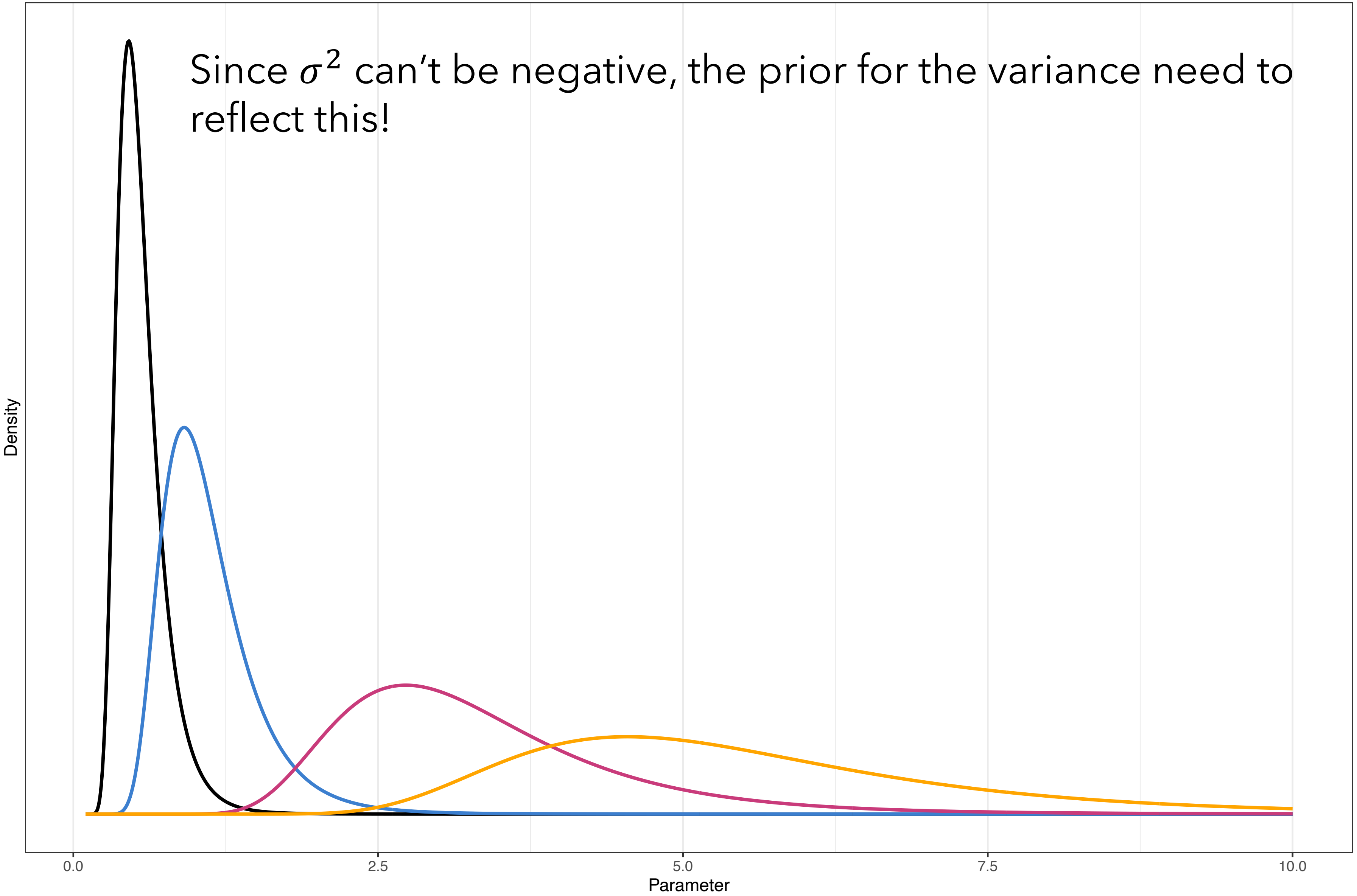


- A good strategy to evaluate the impact of the priors on results is to conduct **sensitivity analysis** where the same model is fit multiple times with different priors

- **We will see this in action in a few minutes**

- Key takeaway is that **you** get to choose the prior and it can be sourced from a variety of places!

Examples of Different Inverse Gamma Priors for σ^2



Prior

$IG(10, 5)$

$IG(10, 10)$

$IG(10, 30)$

$IG(10, 50)$

- The IG distribution is characterized by a shape and scale parameter)
- Today we will use default priors specified by the software

Specifying the Model in R (brms) with Default Priors

- The **brms** (Bürkner, 2017) package provides an interface to fit Bayesian generalized (non-)linear multivariate multilevel model
- It is a wrapper for **Stan**, a popular program that uses MCMC to estimate Bayesian models
- Full code will be available at **https://wpa2024bayesian.ajmquant.com/**
- By default, **brms** uses flat priors for the regression slopes and t distributions for the intercept/SD

```
require(brms) # load the brms package
fit_hci <- brm(
          data = HCI_Data, # the dataset with the four variables
          family = gaussian(),
          hci ~ 1 + SPD + HCSE + HCOE,
          iter = 5000,
          seed = 2024
          )
summary(fit_hci)
```

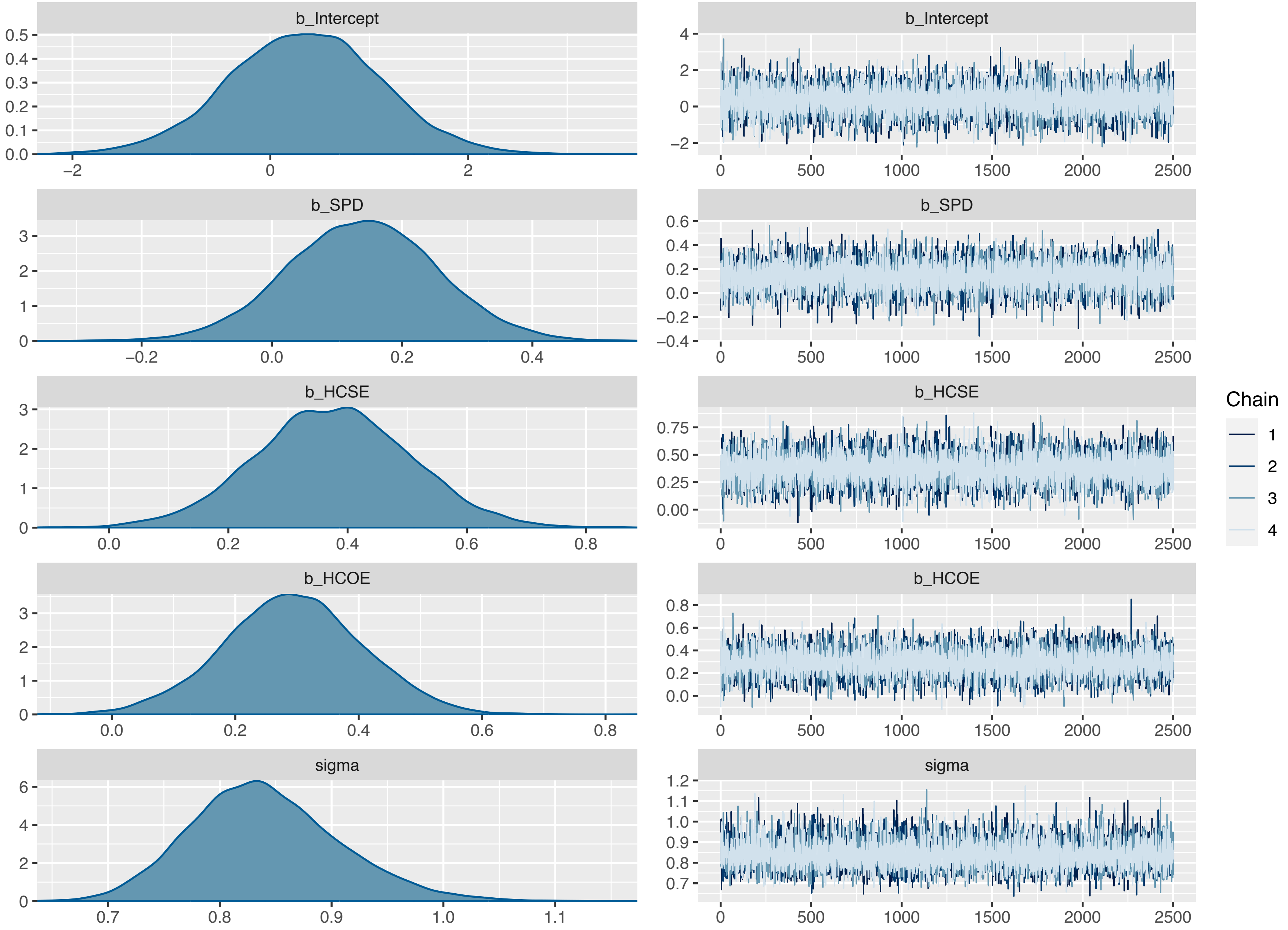
$$\text{HCI}_i = \beta_0 + \beta_1\text{SPD}_i + \beta_2\text{HCSE}_i + \beta_3\text{HCOE}_i + \epsilon_i$$

By default, **brms** implements 4 chains, this may or may not be more than enough depending on your specific model

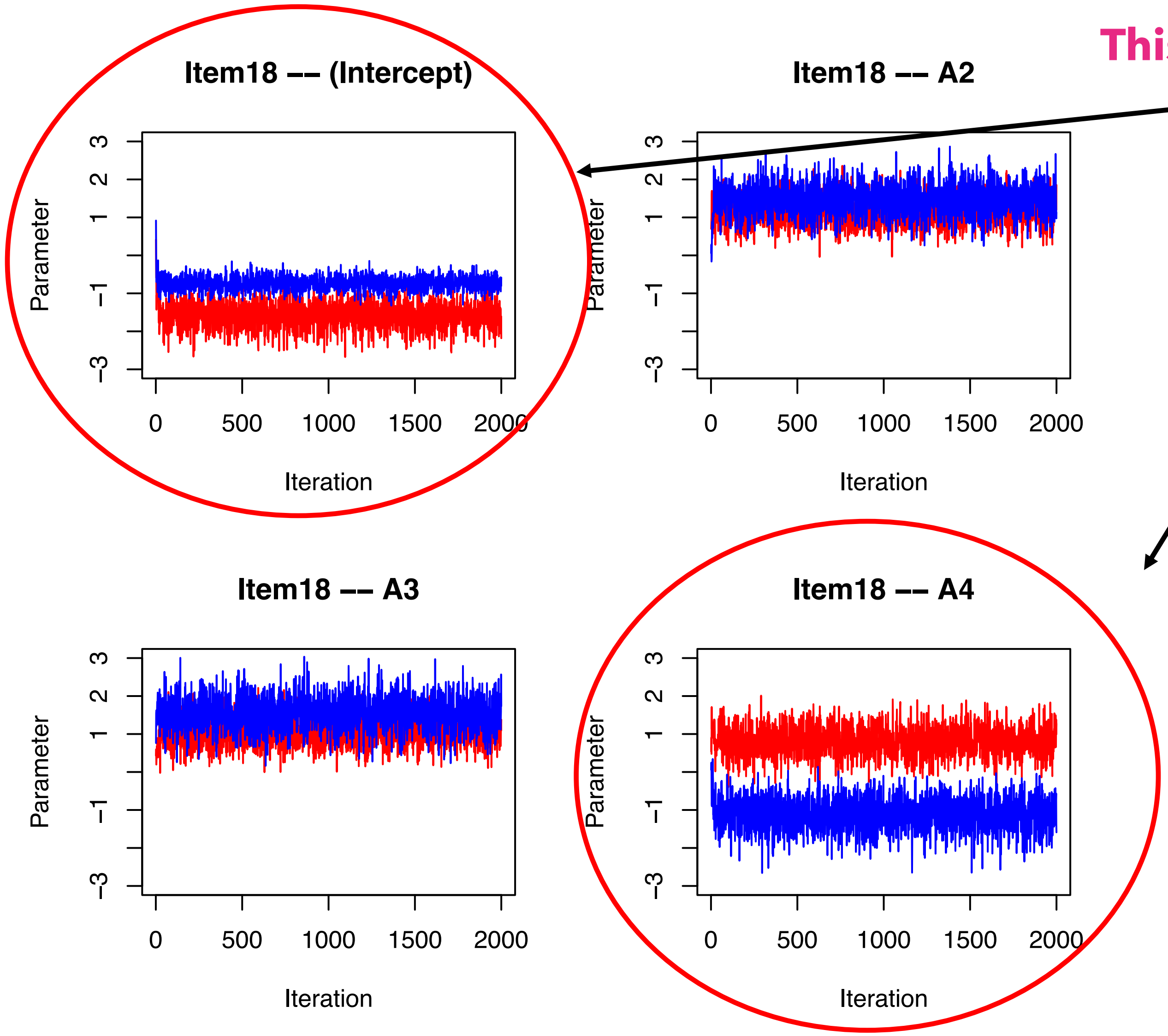
- In the `data` line, we tell R what the name of our dataset is
- The `family` line is used to specify a **linear regression model**
  - Other options include `family = poisson()` for Poisson regression or `family = bernoulli()` for logistic regression, etc.
- The `iter =` line specifies the number of MCMC iterations (by default, half are discarded as burn-in)
- The `seed =` line sets a seed value so we can replicate the analysis and get the same results
- The `summary(fit_hci)` line returns a processed summary of the analysis

Visualizing the Posteriors & Diagnostic Checks

- ❑ Left panel shows the posterior distribution of each parameter
  - ❑ **Everything we want to know about each parameter is contained in its distribution!**
- ❑ Right panel shows the trace plots of the MCMC for each parameter. Looks like we got good mixing!
- ❑ PSRF for each parameter was 1.00 (up to 2 digits of precision) indicating chain convergence
- ❑ Diagnostics look good, so can go ahead and interpret the results
- ❑ BTW, there are more sophisticated ways to check for convergence, this is just the start



Example of Chains that Haven't Converged



This is not what we want to see!

- These issues usually occur when there are multiple modes in the posterior (e.g., mixture models)
- It indicates that the chains are **exploring different parts** of the **parameters space**

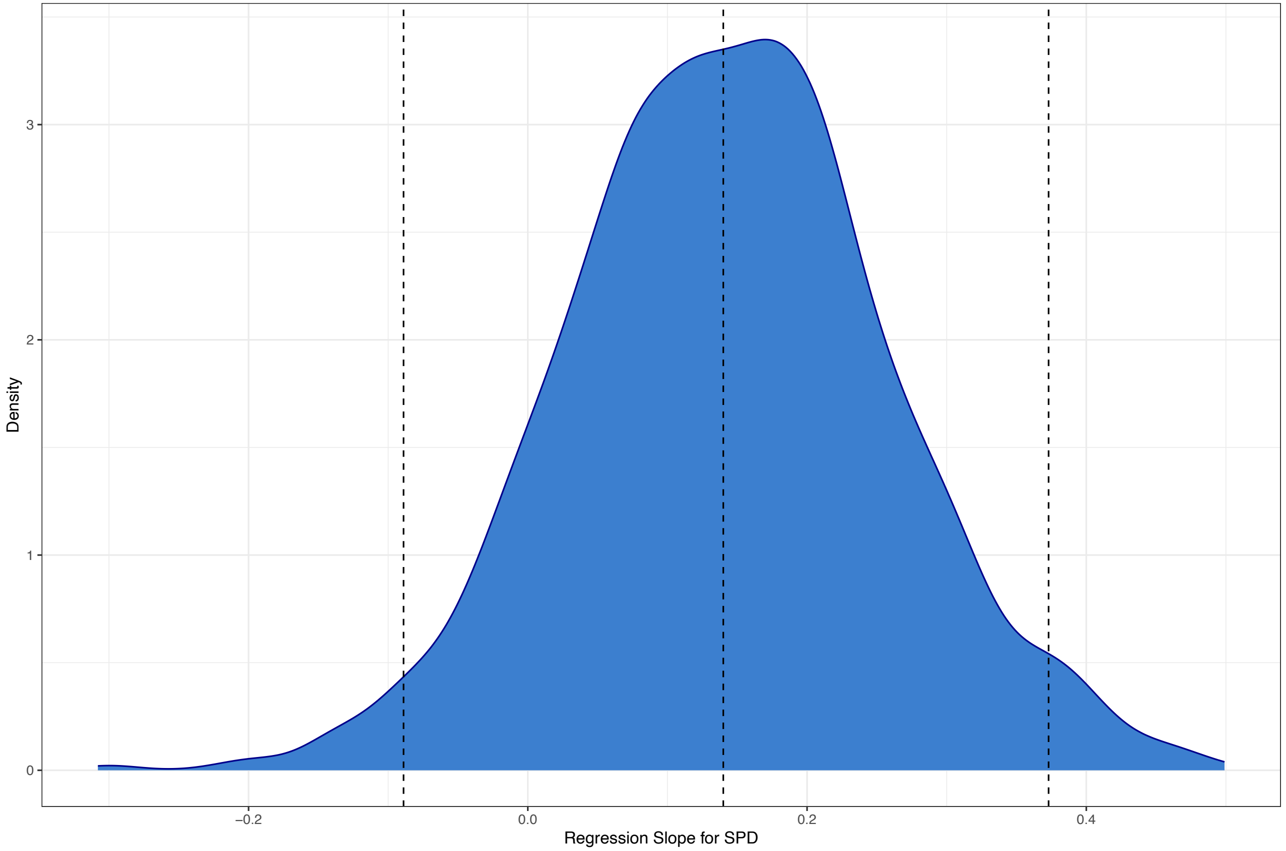
Results and Interpretation

$$\mathrm{HCI}_i = \beta_0 + \beta_1 \mathrm{SPD}_i + \beta_2 \mathrm{HCSE}_i + \beta_3 \mathrm{HCOE}_i + \epsilon_i$$

Bayesian Output with Default Priors (via `brms`)

	Post. Mean	Post. SD	95% Cred. Int.
β_0	0.349	0.772	(-1.186, 1.877)
β_1	0.140	0.116	(-0.089, 0.373)
β_2	0.374	0.130	(0.118, 0.628)
β_3	0.293	0.116	(0.062, 0.519)
σ	0.841	0.066	(0.723, 0.980)

Posterior distribution for β_1



- ❑ Given our data of $n = 85$ middle schoolers, there is there is a 0.95 probability that the effect of SPD on HCI (β_1) is between -0.089 and 0.373 with an **average effect of 0.140**
- ❑ Since this interval contains 0, it is possible that there is not effect of SPD on HCI

Results and Interpretation

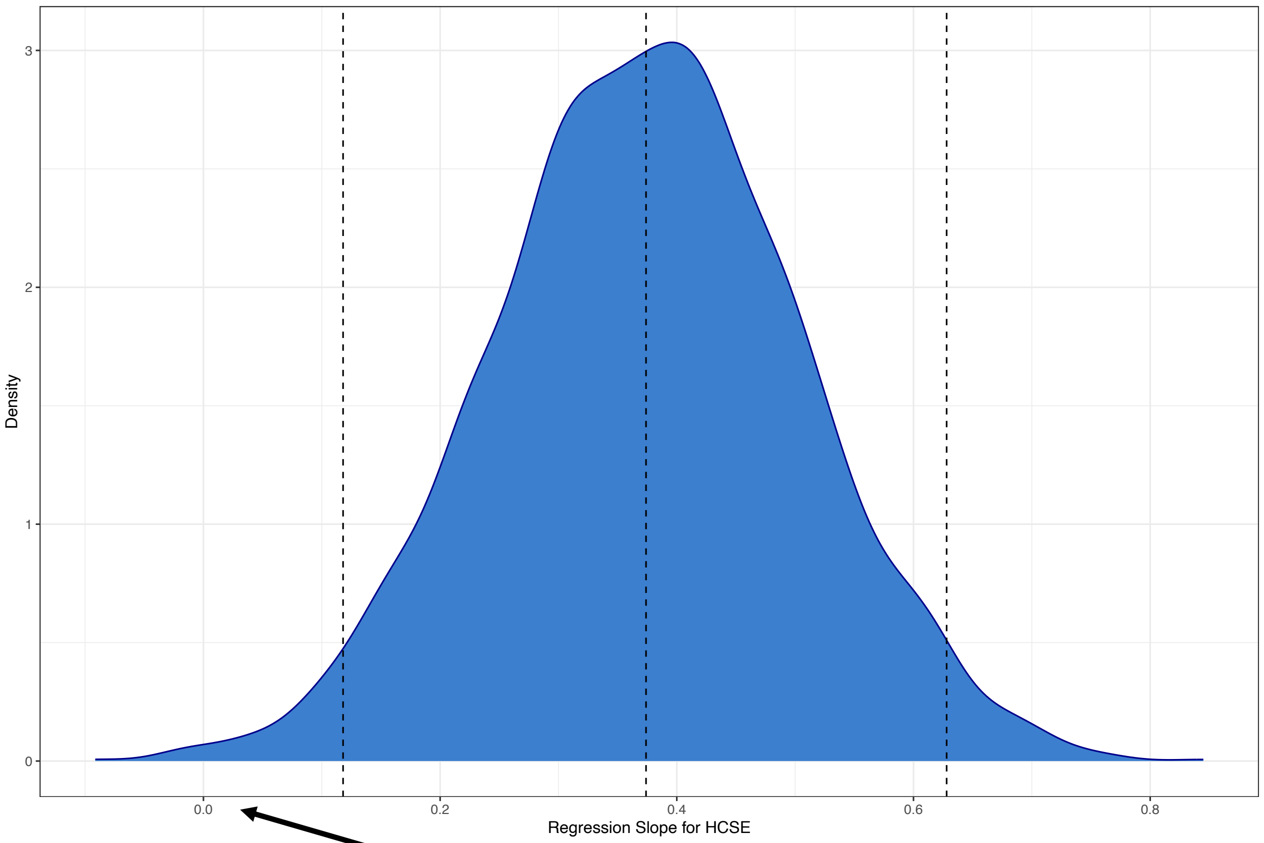
$$HCI_i = \beta_0 + \beta_1 SPD_i + \beta_2 HCSE_i + \beta_3 HCOE_i + \epsilon_i$$

Bayesian Output with Default Priors (via `brms`)

	Post. Mean	Post. SD	95% Cred. Int.
β_0	0.349	0.772	(-1.186, 1.877)
β_1	0.140	0.116	(-0.089, 0.373)
β_2	0.374	0.130	(0.118, 0.628)
β_3	0.293	0.116	(0.062, 0.519)
σ	0.841	0.066	(0.723, 0.980)

- ❑ Given our data, there is there is a 0.95 probability that the effect of HCSE on HCI (β_2) is between 0.118 and 0.628 with an **average effect of 0.374**
- ❑ Interpretation: holding all other variables constant, **a one unit increase in HCSE is associated with a 0.118 to 0.628 unit increase in HCI**

Posterior distribution for β_2



Notice that 0 is not included in the 95% credible interval

Results and Interpretation

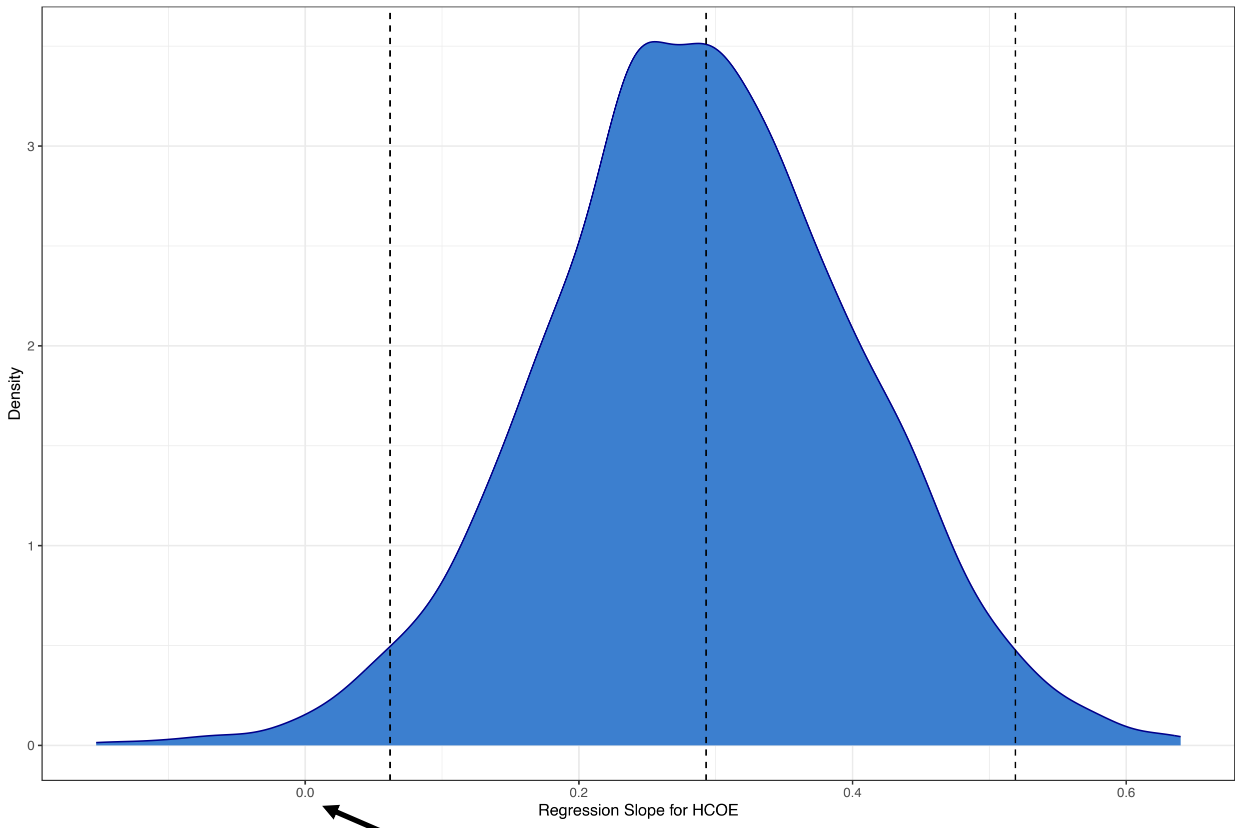
$$\mathrm{HCI}_i = \beta_0 + \beta_1 \mathrm{SPD}_i + \beta_2 \mathrm{HCSE}_i + \beta_3 \mathrm{HCOE}_i + \epsilon_i$$

Bayesian Output with Default Priors (via `brms`)

	Post. Mean	Post. SD	95% Cred. Int.
β_0	0.349	0.772	(-1.186, 1.877)
β_1	0.140	0.116	(-0.089, 0.373)
β_2	0.374	0.130	(0.118, 0.628)
β_3	0.293	0.116	(0.062, 0.519)
σ	0.841	0.066	(0.723, 0.980)

- Given our data, there is there is a 0.95 probability that the effect of HCOE on HCI (β_3) is between 0.062 and 0.519 with an **average effect of 0.293**
- Interpretation: holding all other variables constant, **a one unit increase in HCOE is associated with a 0.062 to 0.519 unit ncrease in HCI**

Posterior distribution for β_3



Notice that 0 is not included in the 95% credible interval

Results and Interpretation

$$\mathrm{HCI}_i = \beta_0 + \beta_1 \mathrm{SPD}_i + \beta_2 \mathrm{HCSE}_i + \beta_3 \mathrm{HCOE}_i + \epsilon_i$$

Bayesian Output with Default Priors (via `brms`)

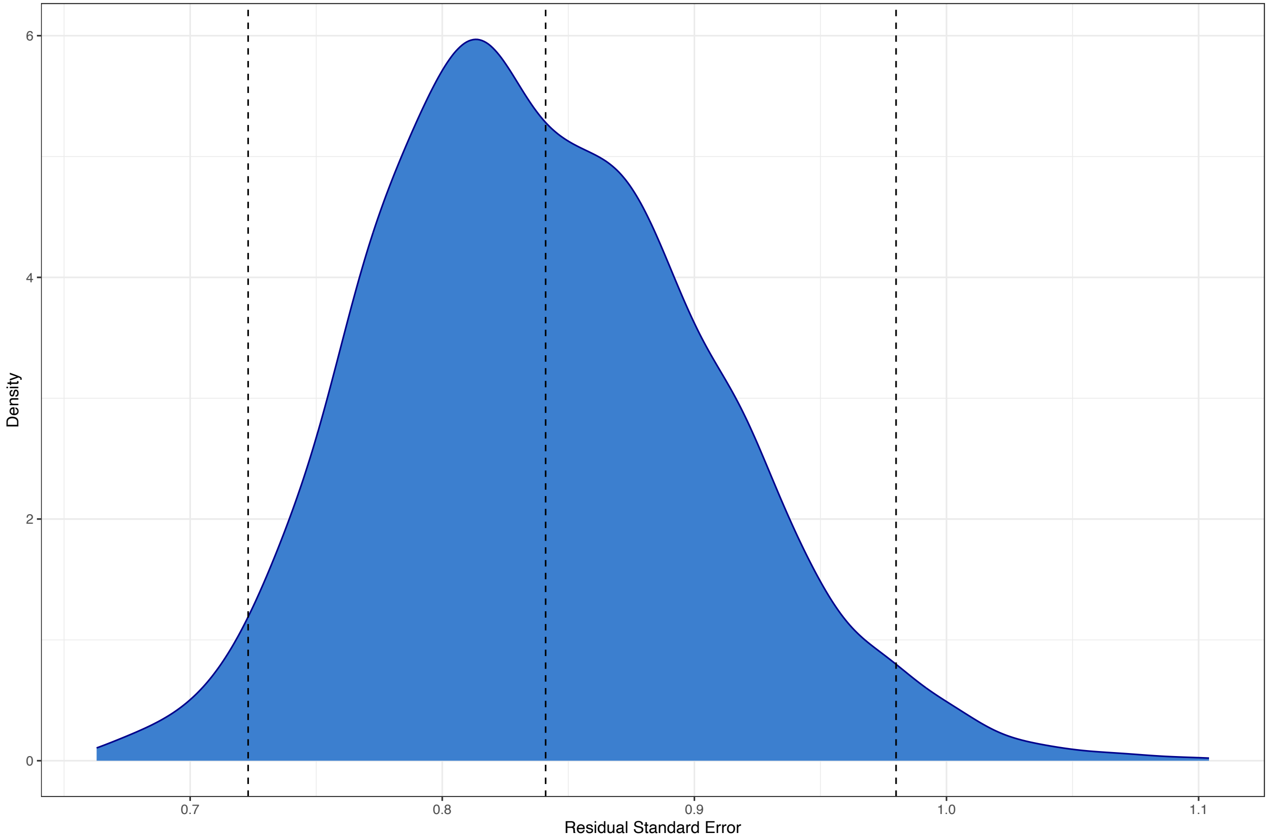
	Post. Mean	Post. SD	95% Cred. Int.
β_0	0.349	0.772	(-1.186, 1.877)
β_1	0.140	0.116	(-0.089, 0.373)
β_2	0.374	0.130	(0.118, 0.628)
β_3	0.293	0.116	(0.062, 0.519)
σ	0.841	0.066	(0.723, 0.980)

- Because σ is treated as a random variable, it also has a posterior distribution

- Uncertainty about the uncertainty

- Here, there is a 0.95 probability the population σ is between 0.723 and 0.980

Posterior distribution for σ



Quick Aside: Least Squares vs Bayesian Analysis

Least Squares Estimation (via `lm`)			
	Estimate	SE	95% Conf. Int.
β_0	0.361	0.772	(-1.155, 1.876)
β_1	0.140	0.116	(-0.087, 0.366)
β_2	0.374	0.130	(0.116, 0.631)
β_3	0.291	0.116	(0.068, 0.514)
σ	0.829		

Bayesian Output with Default Priors (via `brms`)			
	Post. Mean	Post. SD	95% Cred. Int.
β_0	0.349	0.772	(-1.186, 1.877)
β_1	0.140	0.116	(-0.089, 0.373)
β_2	0.374	0.130	(0.118, 0.628)
β_3	0.293	0.116	(0.062, 0.519)
σ	0.841	0.066	(0.723, 0.980)

Results are numerically similar but conceptually different!

Least Squares

- The β estimates are the values that **minimize** the residual sum of squares
- The standard errors of β is refers the **sampling distribution** of β under **repeated sampling**
- Under repeated sampling, 95% of CIs would contain the true population β values
- No way to incorporate your domain expertise into the analysis

Bayesian

- Given **the observed data**, the probability that β between (L, U) is 0.95 → $L \leq \beta \leq U$ is a 95% credible interval
- The posterior SD is the SD of the posterior distribution of β, not that of a hypothetical sampling distribution
- You have control over how much influence you incorporate into analysis

Specifying the Model in R (`brms`) with User-Defined Priors

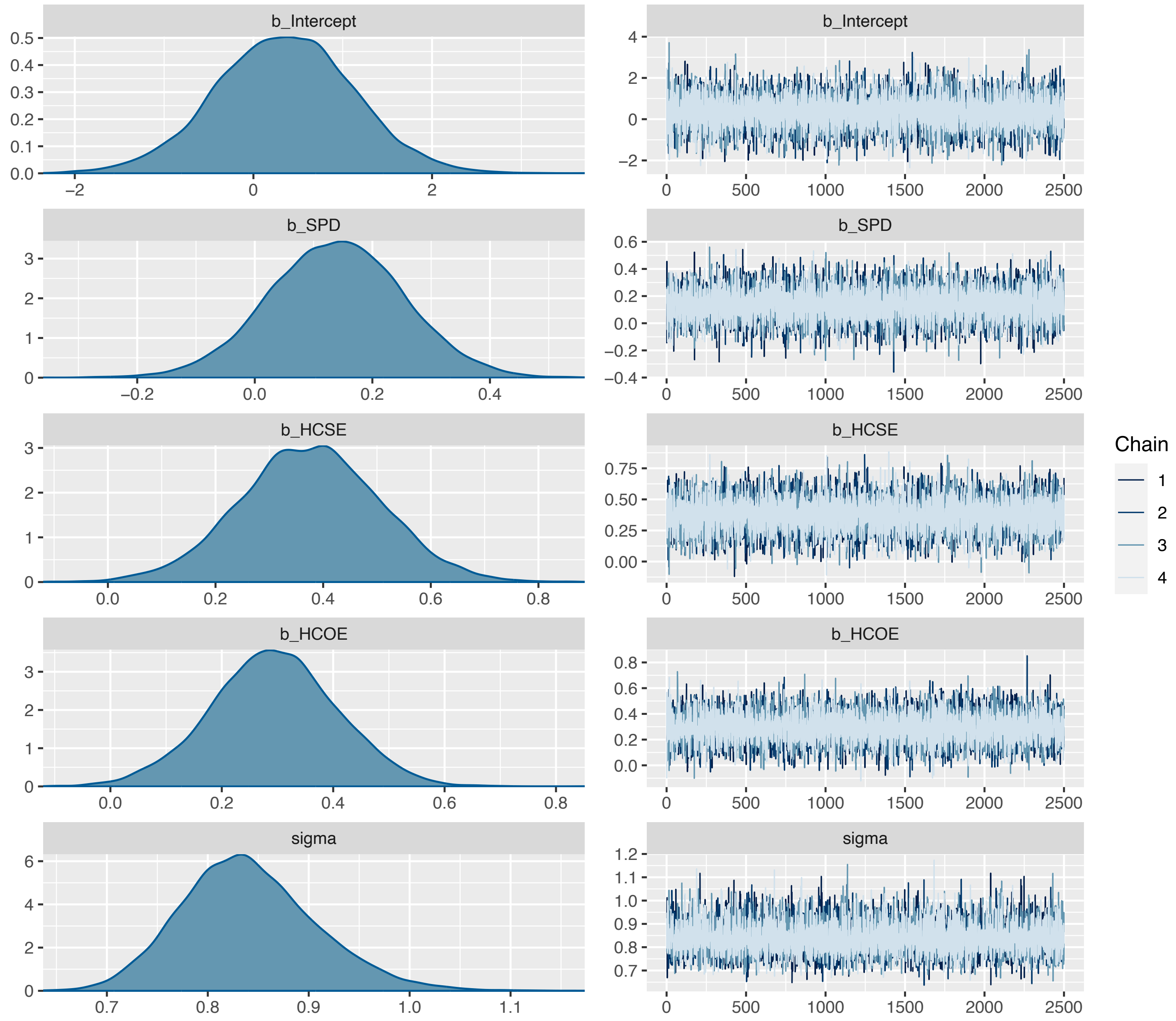
- The first version of the model used the default flat priors (uninformative) so the data "did most of the talking"
- How (if at all) do the results change if we start implement different priors?
- **Sensitivity analysis**

```
require(brms)
fit_hci <- brm(
          data = HCI_Data,
          family = gaussian(),
          hci ~ 1 + SPD + HCSE + HCOE,
          prior = c(
                  prior_string("normal(0, 10)", class = "Intercept"),
                  prior_string("normal(0, 10)", class = "b")
            ),
          iter = 5000,
          seed = 2024
          )
summary(fit_hci)
```

$$\mathrm{HCI}_i = \beta_0 + \beta_1 \mathrm{SPD}_i + \beta_2 \mathrm{HCSE}_i + \beta_3 \mathrm{HCOE}_i + \epsilon_i$$

Note, `brms` uses SDs instead of variances so here we are assigning $N(0, 100)$ priors to the intercept and regression slopes

Visualizing the Posteriors & Diagnostic Checks



Comparing Default Priors to $N(0, 10)$ Priors

$$\text{HCI}_i = \beta_0 + \beta_1 \text{SPD}_i + \beta_2 \text{HCSE}_i + \beta_3 \text{HCOE}_i + \epsilon_i$$

Bayesian Output with Default Priors (via brms)

	Post. Mean	Post. SD	95% Cred. Int.
β_0	0.349	0.772	(-1.186, 1.877)
β_1	0.140	0.116	(-0.089, 0.373)
β_2	0.374	0.130	(0.118, 0.628)
β_3	0.293	0.116	(0.062, 0.519)
σ	0.841	0.066	(0.723, 0.980)

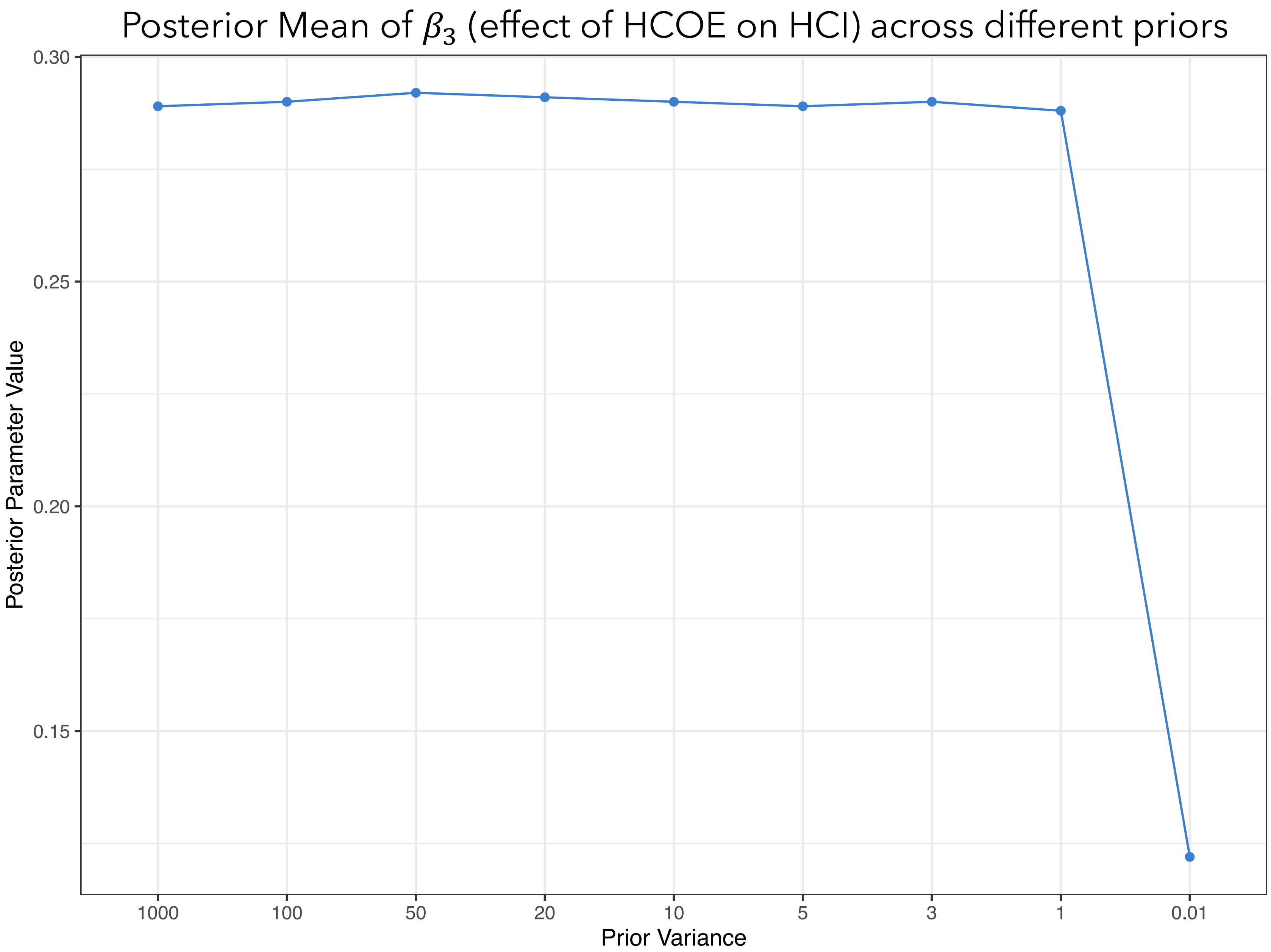
Bayesian Output with $N(0, 10)$ Priors (via brms)

	Post. Mean	Post. SD	95% Cred. Int.
β_0	0.346	0.769	(-1.178, 1.860)
β_1	0.139	0.115	(-0.087, 0.365)
β_2	0.376	0.131	(0.119, 0.635)
β_3	0.294	0.113	(0.068, 0.516)
σ	0.842	0.067	(0.722, 0.984)

Essentially the same!

- This is not too surprising given that a $N(0, 10)$ prior is relatively large given the magnitudes of the effects
- What if we use other more concentrated priors?
- Redid analyses with the following priors: $N(0, v)$ where $v \in \{1000, 100, 50, 20, 10, 5, 3, 1, 0.1\}$

Impact of Priors on Posterior Mean



- The priors have **virtually no influence** on the results for more variance value
- This suggests the results are robust to different prior choices
- The only exception is the $N(0, 0.1)$ prior which **pulled the parameter estimate towards 0**
  - The
- Similar patterns were observed for the other parameters

Sequential Posterior Updating / Accumulation of Evidence

- We can continue to update our beliefs by treating the posterior as a prior in a future analysis!

Analysis 1 (What we've been doing so far)

$$\widetilde{P(\theta)} = P(\theta \mid y_1) \propto P(y_1 \mid \theta)P(\theta)$$

Updating our beliefs about parameters given the observed data (y_1 is the dataset we've been analyzing up until now)

Analysis 2 (accumulation of evidence)

- "Yesterdays posterior is tomorrow's prior"

$$P(\theta \mid y_2, y_1) \propto P(y_2 \mid \theta)\widetilde{P(\theta)}$$

Continuously updating our beliefs when we have new information available (y_2 is a new dataset that is collected at some point in the future)

A technical caveat

- Technically, the above assumes conditional independence between y_1 and y_2 given θ)
  - In other words, θ completely describes the data generating process

$$P(\theta \mid y_1, y_2) \propto P(y_2 \mid \theta)P(y_1 \mid \theta)P(\theta)$$

Sequential Posterior Updating / Accumulation of Evidence

$$\mathrm{HCI}_i = \beta_0 + \beta_1 \mathrm{SPD}_i + \beta_2 \mathrm{HCSE}_i + \beta_3 \mathrm{HCOE}_i + \epsilon_i$$

- Results from the previous analysis form the priors of the new analysis (empirical priors)

Bayesian Output with Default Priors (via `brms`)

	Post. Mean	Post. SD	95% Cred. Int.
β_0	0.349	0.772	(-1.186, 1.877)
β_1	0.140	0.116	(-0.089, 0.373)
β_2	0.374	0.130	(0.118, 0.628)
β_3	0.293	0.116	(0.062, 0.519)
σ	0.841	0.066	(0.723, 0.980)

```
fit_hci_new <- brm(data = HCI_Data_New,
                   family = gaussian(),
                   hci ~ 1 + SPD + HCSE + HCOE,
                   prior = c(
                     set_prior("normal(0.349, 0.772)", class = "Intercept"),
                     set_prior("normal(0.140, 0.116)", class = "b", coef = "SPD"),
                     set_prior("normal(0.374, 0.130)", class = "b", coef = "HCSE"),
                     set_prior("normal(0.293, 0.116)", class = "b", coef = "HCOE")
                   ),
                   iter = 5000, seed = 2025)
```

New dataset from same population

Sequential Posterior Updating / Accumulation of Evidence

$$\text{HCI}_i = \beta_0 + \beta_1 \text{SPD}_i + \beta_2 \text{HCSE}_i + \beta_3 \text{HCOE}_i + \epsilon_i$$

Original Analysis

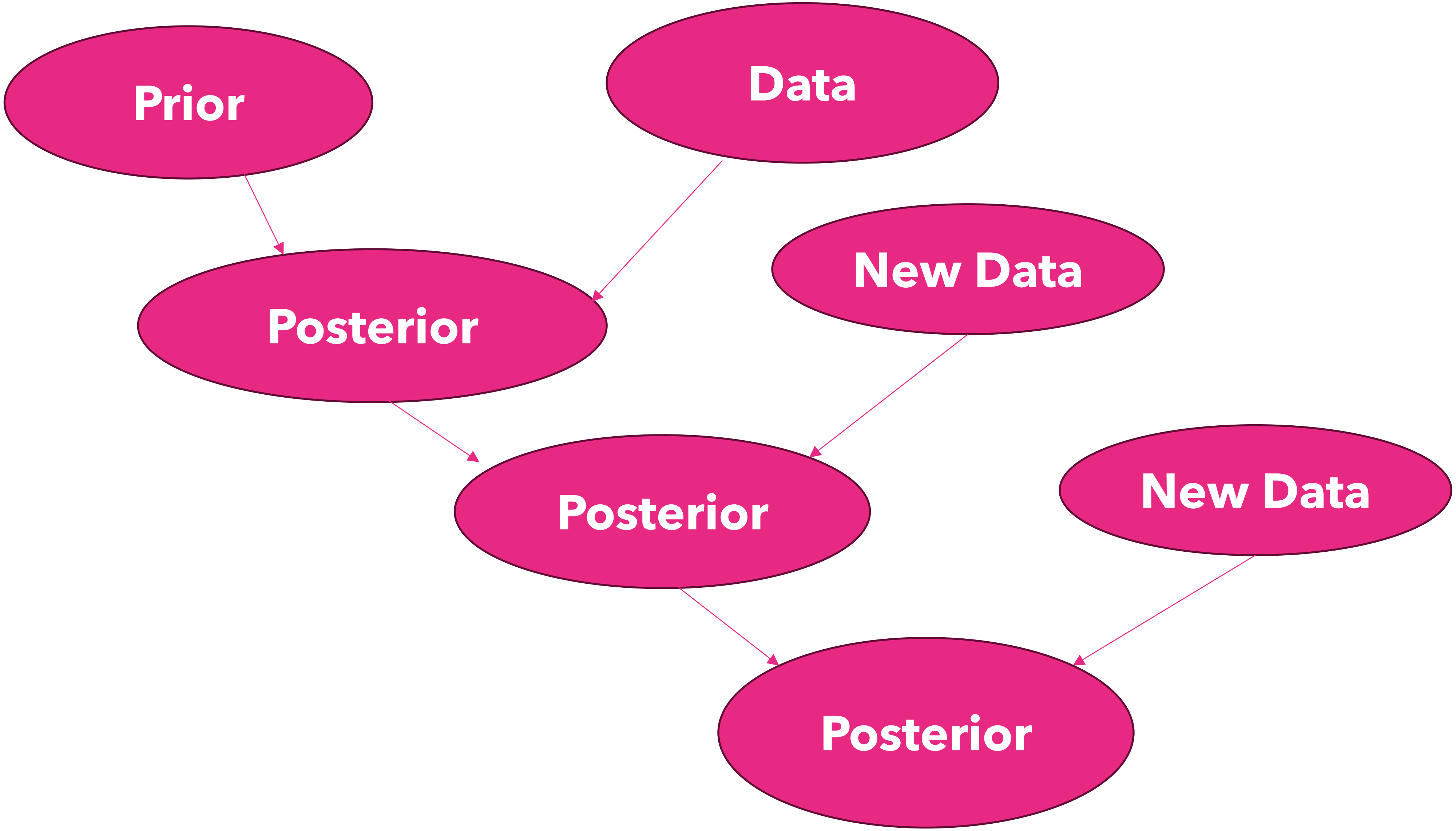
Bayesian Output with Default Priors (via `brms`)			
	Post. Mean	Post. SD	95% Cred. Int.
β_0	0.349	0.772	(-1.186, 1.877)
β_1	0.140	0.116	(-0.089, 0.373)
β_2	0.374	0.130	(0.118, 0.628)
β_3	0.293	0.116	(0.062, 0.519)
σ	0.841	0.066	(0.723, 0.980)

Analysis on New Data

Bayesian Output with Posterior Priors (via `brms`)			
	Post. Mean	Post. SD	95% Cred. Int.
β_0	0.278	0.513	(-0.728, 1.282)
β_1	0.229	0.080	(0.070, 0.386)
β_2	0.326	0.091	(0.145, 0.505)
β_3	0.285	0.074	(0.141, 0.429)
σ	0.817	0.065	(0.701, 0.957)

- ❑ In general, notice how the Posterior SDs and credible intervals are narrower
  - ❑ **Our uncertainty about the effects decreases as we collect and analyze new data!**
- ❑ The credible interval for β_1 in the new analysis doesn't contain 0
  - ❑ Evidence of a positive effect of SPD on HCI!

Sequential Posterior Updating / Accumulation of Evidence



Wrapping Up

- Bayesian statistics offers a flexible approach to modeling psychological data
  - Construction
  - Estimation
  - Model fit, model comparisons, etc.

- Incorporating **substantive** and **domain** expertise into an analysis in a principled way

- Use of Bayesian statistics is becoming increasingly popular as packages like **`brms`** make it easy to estimate many models

- Hopefully today's workshop highlighted the usefulness of the Bayesian approach and motivated you to want to learn more!

- Reminder: slides/code will be available at **https://wpa2024bayesian.ajmquant.com/**

Thank you!

Hyeri Hong, Ph.D.

Email: hyerihong@mail.fresnostate.edu

Alfonso J. Martinez

Email: alfonso-martinez@uiowa.edu

Please reach out if you have any questions!