

# **MODULE 3**

## **ESTIMATION AND INFERENCE**

# ESTIMATION OVERVIEW

---

- Maximum likelihood has been the go-to estimator for multilevel models for many years
- MCMC is becoming increasingly popular because of its flexibility in estimating complex MLMs
- Recent methodological innovations (e.g., dynamic SEMs, intensive repeated measures) primarily leverage MCMC

# OUTLINE

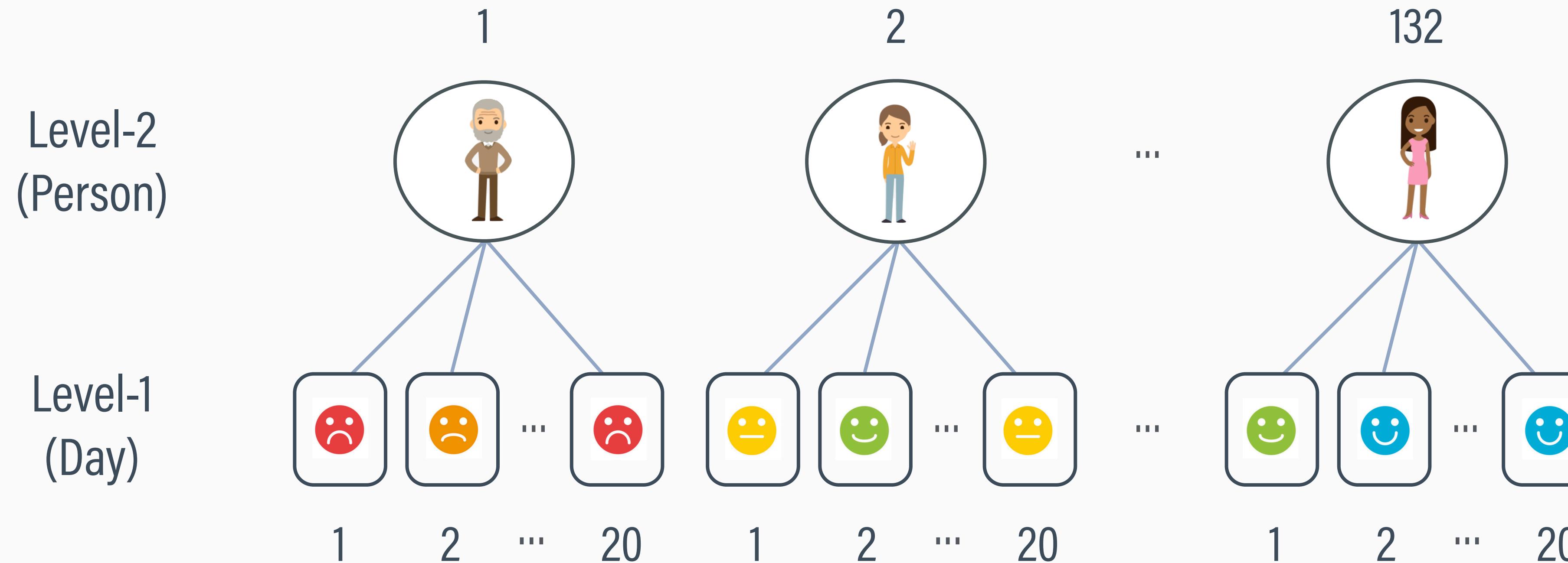
- 1 Analysis Example
- 2 Frequentist vs. Bayesian Statistical Paradigms
- 3 Maximum Likelihood Estimation
- 4 MCMC Estimation
- 5 MCMC Diagnostics
- 6 Checking MLM Assumptions

# OUTLINE

- 1 Analysis Example
- 2 Frequentist vs. Bayesian Statistical Paradigms
- 3 Maximum Likelihood Estimation
- 4 MCMC Estimation
- 5 MCMC Diagnostics
- 6 Checking MLM Assumptions

# DAILY DIARY APPLICATION

- $n_j = 20$  daily positive affect and sleep assessments nested within  $J = 132$  chronic pain patients ( $N = 2680$  data records)



# DATA STRUCTURE

- Data in stacked or long format
- Each level-2 unit (person) has multiple rows, one per level-1 (daily) observation
- The  $i$  subscript indexes level-1 observations, and  $j$  indexes level-2 units

Row	i	j	PAFFECT <sub>ij</sub>	SLEEP <sub>ij</sub>
1	1	1	7.3	5.6
2	2	1	2.5	4.3
...	...	1	...	...
20	20	1	6.3	7.3
21	1	2	4.0	3.9
22	2	2	4.0	7.1
...	...	2	...	...
40	20	2	4.4	3.5
...	...	...	...	...
2621	1	132	3.3	5.4
2622	2	132	4.8	3.5
...	...	132	...	...
2640	20	132	4.8	7.9

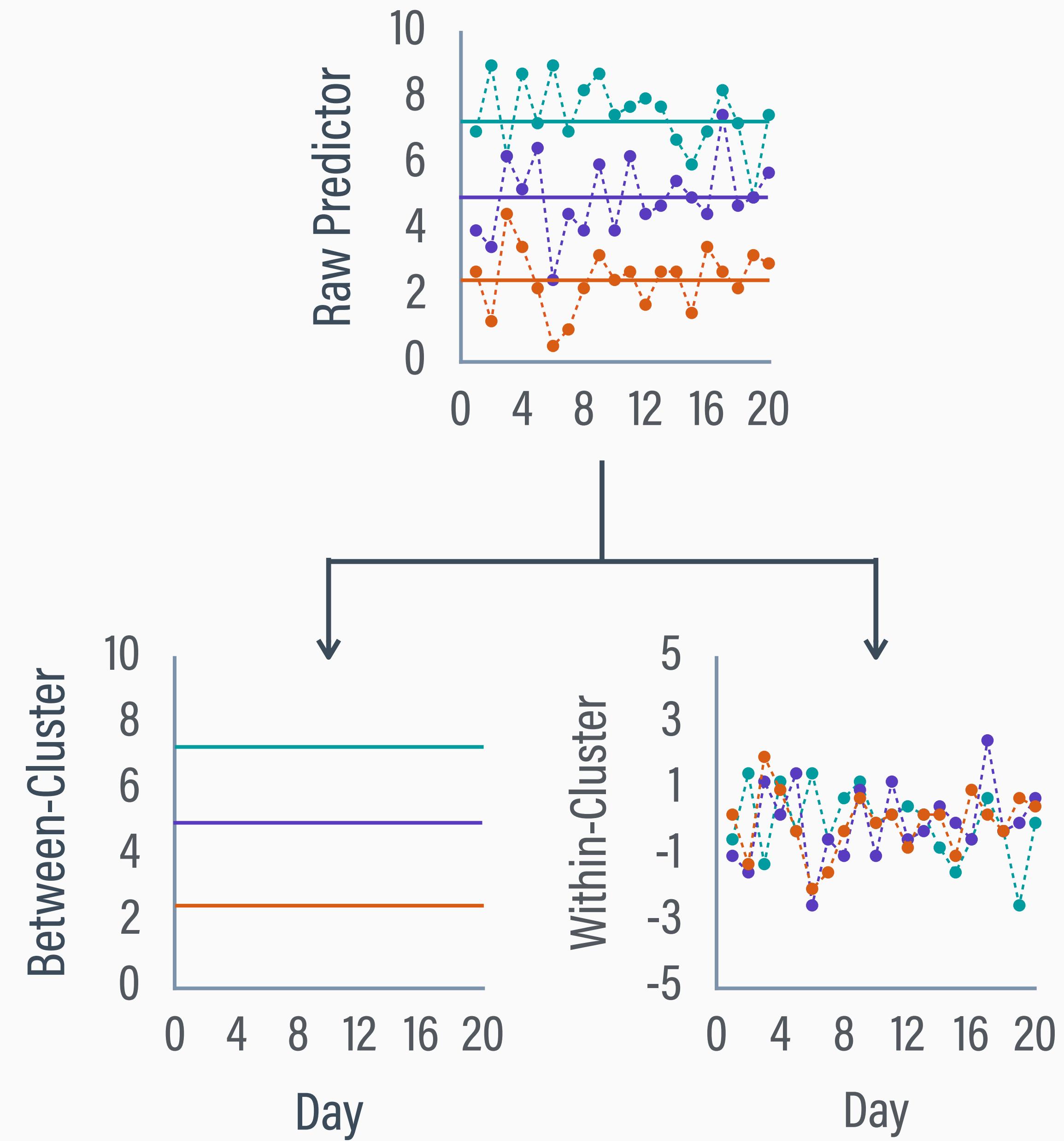
# DISAGGREGATED PREDICTOR

- Disaggregation centers each daily score around its level-2 person mean

$$\text{sleep}^b_j = \mu_j(\text{sleep})$$

$$\text{sleep}^w_{ij} = \text{sleep}_{ij} - \mu_j(\text{sleep})$$

- $\text{sleep}^w$  contains only intraindividual (level-1) variation, and  $\text{sleep}^b$  reflects only person (level-2) mean differences

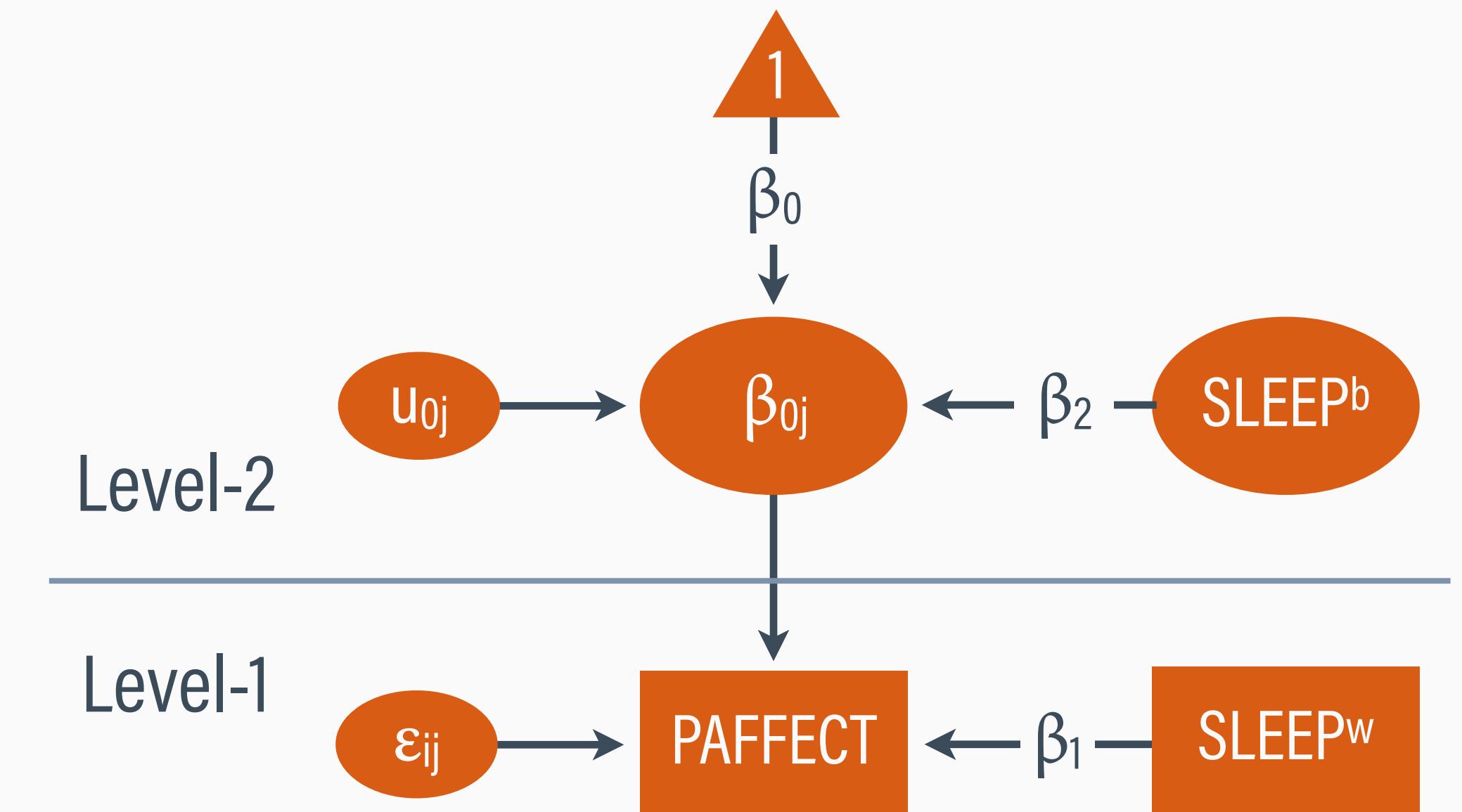


# ANALYSIS MODEL

- Model featuring sleep disaggregated into unique level-1 and level-2 predictors

$$PAFFECT_{ij} = \beta_0 + \beta_1(SLEEP_{ij}^W) + \beta_2(SLEEP_j^B) + u_{0j} + \varepsilon_{ij}$$

- Person-average sleep quality predicts person-average positive mood ( $\beta_2$ ), and within-person fluctuations in daily sleep predict within-person fluctuations in daily affect ( $\beta_1$ )



# BLIMP STUDIO SCRIPT 3.1

---

**DATA:** PainDiary.dat;

**VARIABLES:** Person Day PosAffect NegAffect Pain WorkGoal LifeGoal SleepQual Female Education  
Employment MarStatus NumDiagnose ActivityLevel PainAccept Catastrophize Stress Anxiety;

**CLUSTERID:** Person;

**CENTER:**

grandmean = SleepQual.mean; # defines the fixed (average) intercept as the grand mean

groupmean = SleepQual; # cwc with level-2 latent group means

**MODEL:** PosAffect ~ intercept SleepQual SleepQual.mean | intercept; # .mean invokes latent means

**BURN:** 10000;

**ITERATIONS:** 10000;

**SEED:** 90291;

# RBLIMP SCRIPT 3 (MODEL 1)

---

```
model1 <- rblimp(  
  data = PainDiary,  
  clusterid = 'Person',  
  center = 'grandmean = SleepQual.mean; groupmean = SleepQual',  
  model = 'PosAffect ~ intercept SleepQual SleepQual.mean | intercept',  
  seed = 90291,  
  burn = 10000,  
  iter = 10000)  
  
output(model1)
```

# BLIMP OUTPUT

□ = level-2 estimate

□ = level-1 estimate

Outcome Variable: PosAffect

Group Mean Centered: SleepQual

Parameters	Estimate	StdDev	2.5%	97.5%	ChiSq	PValue	N_Eff
<hr/>							
Variances:							
L2 : Var(Intercept)	<b>1.839</b>	0.246	1.436	2.399	---	---	5051.168
Residual Var.	<b>1.307</b>	0.037	1.238	1.380	---	---	8227.580
<hr/>							
Coefficients:							
Intercept	<b>4.997</b>	0.133	4.761	5.277	1423.604	0.000	154.295
SleepQual	<b>0.173</b>	0.013	0.147	0.199	174.595	0.000	9302.345
SleepQual.mean[Person]	<b>0.598</b>	0.094	0.416	0.788	40.936	0.000	161.345
<hr/>							
Standard Deviations:							
L2 : SD(Intercept)	<b>1.356</b>	0.090	1.198	1.549	---	---	5041.908
Residual SD	<b>1.143</b>	0.016	1.113	1.175	---	---	8223.888
<hr/>							
Standardized Coefficients:							
SleepQual	0.151	0.013	0.127	0.176	142.389	0.000	1996.211
SleepQual.mean[Person]	0.412	0.055	0.297	0.513	55.809	0.000	169.087
<hr/>							
Proportion Variance Explained							
by Coefficients	0.193	0.044	0.112	0.286	---	---	175.664
by Level-2 Random Intercepts	0.471	0.042	0.392	0.555	---	---	414.230
by Level-1 Residual Variation	0.334	0.027	0.280	0.387	---	---	511.829

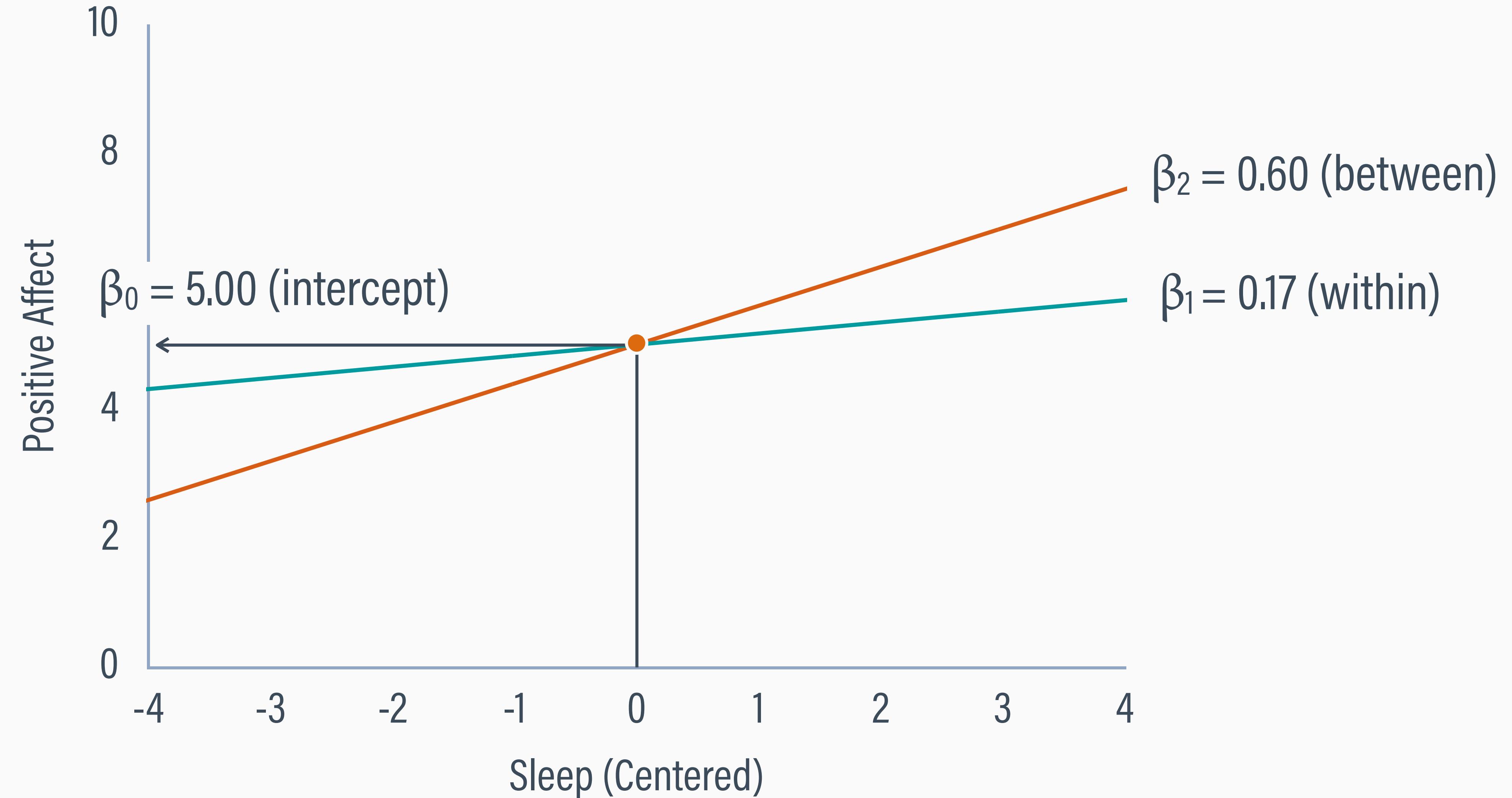
# FIXED EFFECT INTERPRETATIONS

---

- $\beta_0 = 5.00$  is the positive affect grand mean (because both predictors are centered)
- $\beta_1 = 0.17$  is the expected affect difference between two daily sleep scores from the same person that differ by one point
- $\beta_2 = 0.60$  is the expected affect difference between two individuals whose average sleep ratings differ by one point

# LEVEL-1 AND LEVEL-2 REGRESSIONS

---



# RANDOM EFFECT INTERPRETATIONS

---

- $u_{0j} = \beta_{0j} - (\beta_0 + \beta_2(\text{sleep}_j^b))$
- $\text{var}(u_{0j}) = 1.84$  is the average squared distance between the level-2 affect means and their predicted values
- $\text{sd}(u_{0j}) = 1.36$  is the average distance between the level-2 affect means and their predicted values
- $\varepsilon_{ij} = \text{pffect}_{ij} - (\beta_{0j} + \beta_1(\text{sleep}_{ij}^w))$
- $\text{var}(\varepsilon_{ij}) = 1.31$  is the average squared distance between the level-1 affect observations and their predicted values
- $\text{sd}(\varepsilon_{ij}) = 1.14$  is the average distance between the level-1 affect observations and their predicted values

# OUTLINE

- 1 Analysis Example
- 2 Frequentist vs. Bayesian Statistical Paradigms
- 3 Maximum Likelihood Estimation
- 4 MCMC Estimation
- 5 MCMC Diagnostics
- 6 Checking MLM Assumptions

# FREQUENTIST STATISTICAL PARADIGM

---

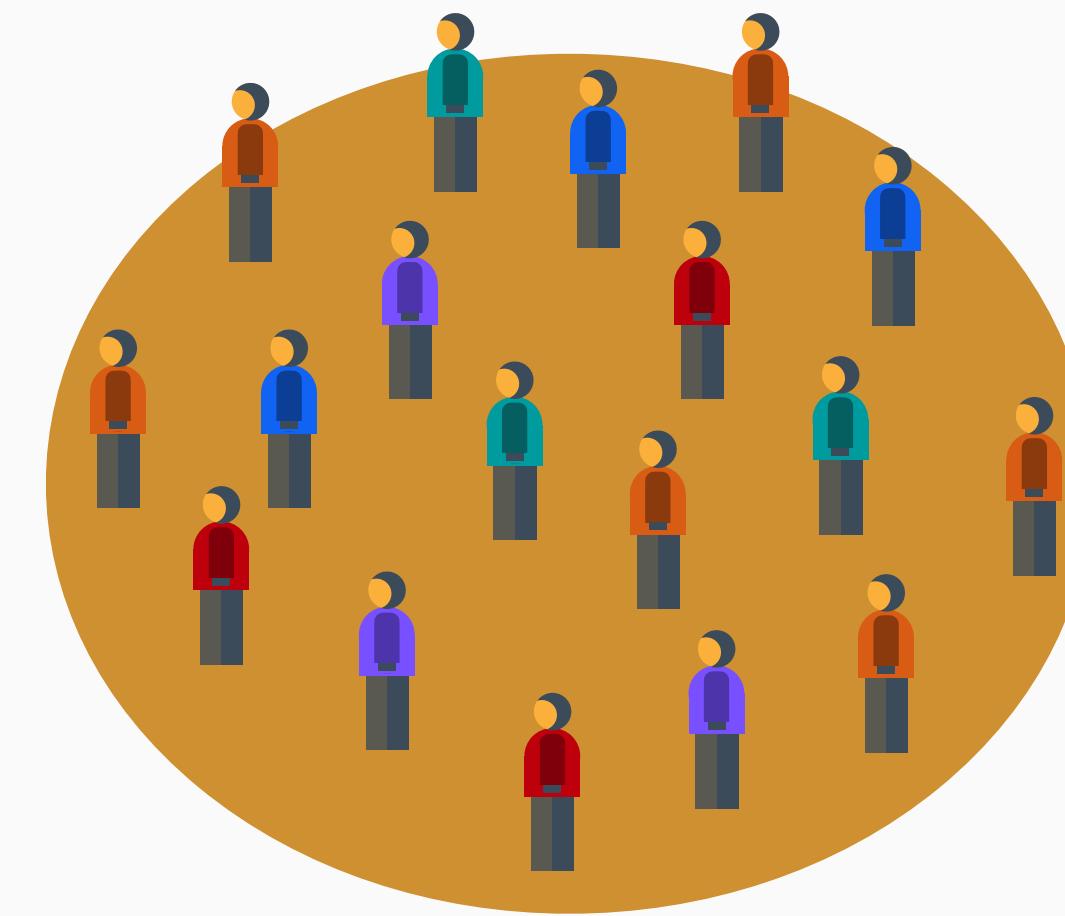
- The frequentist paradigm is defined by the idea that there is one population with unknown parameters (e.g.,  $\mu$  and  $\sigma$ )
- We imagine numerous hypothetical samples of size  $N$  from the population, each with its own unique estimates (e.g.,  $\bar{X}$  and  $s$ )
- The population-level statistics (parameters) are locked in at a single set of values, whereas the sample-level statistics (estimates) vary across different data sets

# FREQUENTIST FRAMEWORK

---

One population with unknown parameters  $\beta_0$ ,  $\beta_1$ , and  $\beta_2$

$$\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)$$

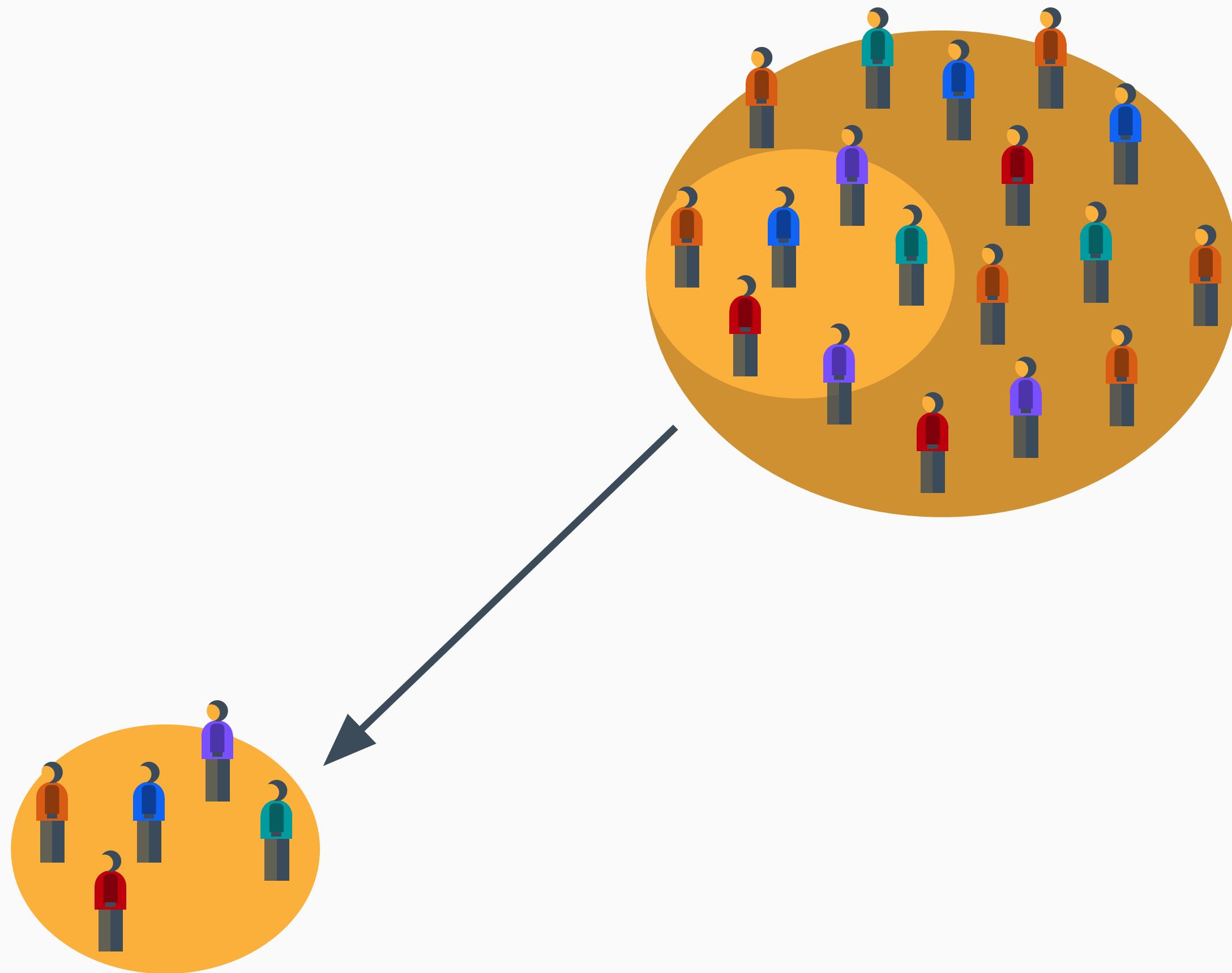


# FREQUENTIST FRAMEWORK

---

One population with unknown parameters. Estimates vary across different hypothetical samples.

$$\beta = (\beta_0, \beta_1, \beta_2)$$



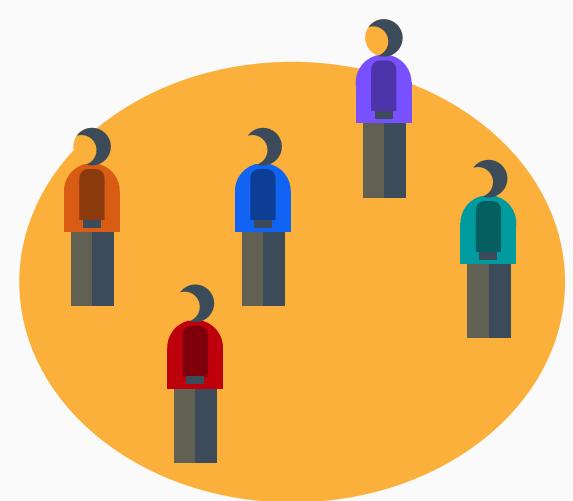
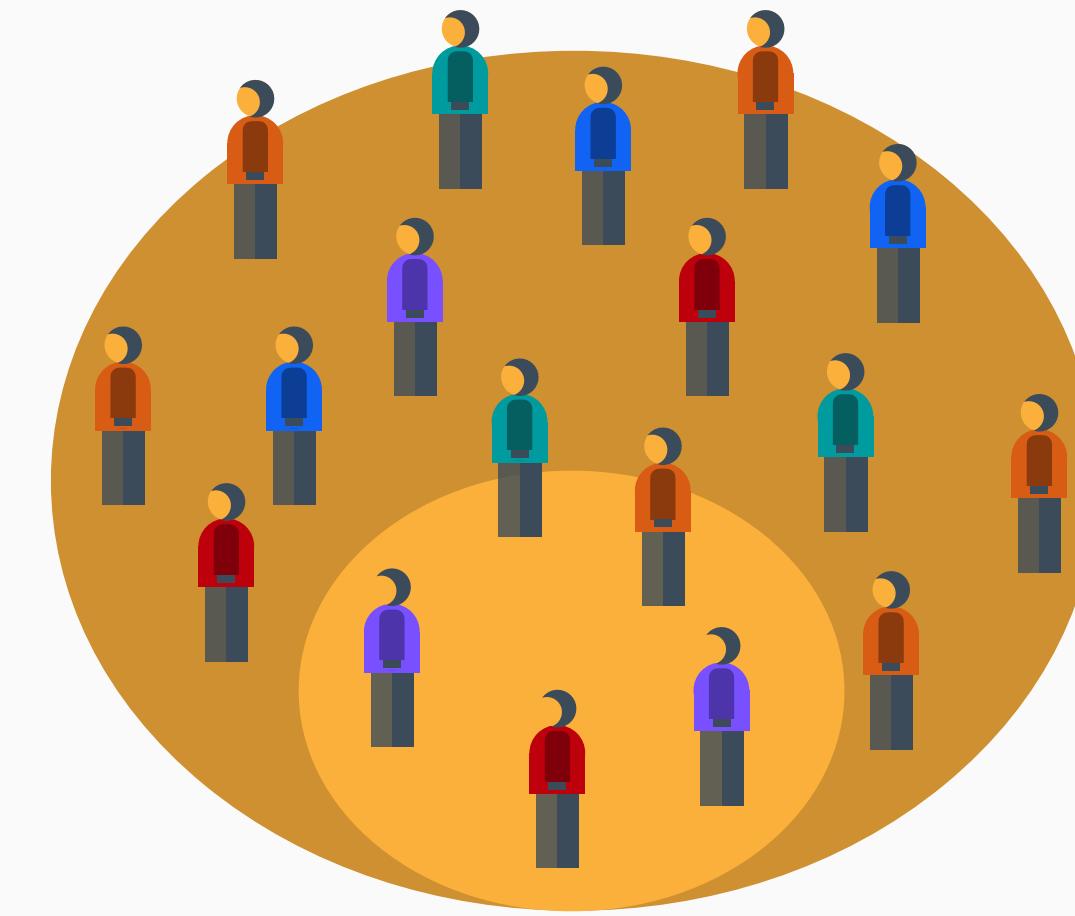
$$\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2)$$

# FREQUENTIST FRAMEWORK

---

One population with unknown parameters. Estimates vary across different hypothetical samples.

$$\beta = (\beta_0, \beta_1, \beta_2)$$



$$\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2)$$

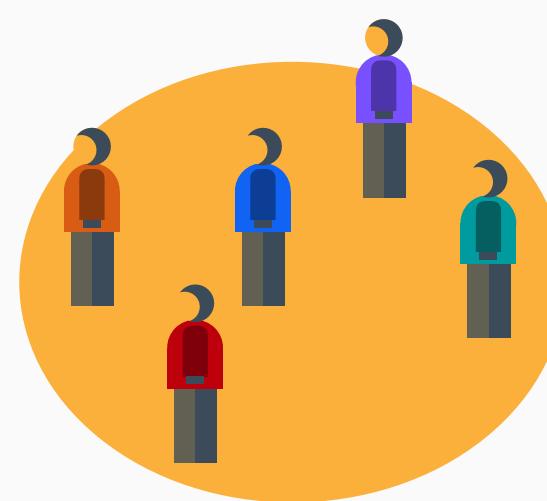


$$\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2)$$

# FREQUENTIST FRAMEWORK

One population with unknown parameters. Estimates vary across different hypothetical samples.

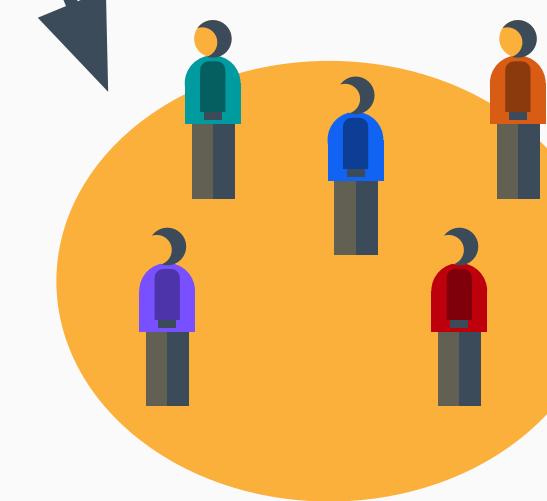
$$\beta = (\beta_0, \beta_1, \beta_2)$$



$$\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2)$$



$$\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2)$$

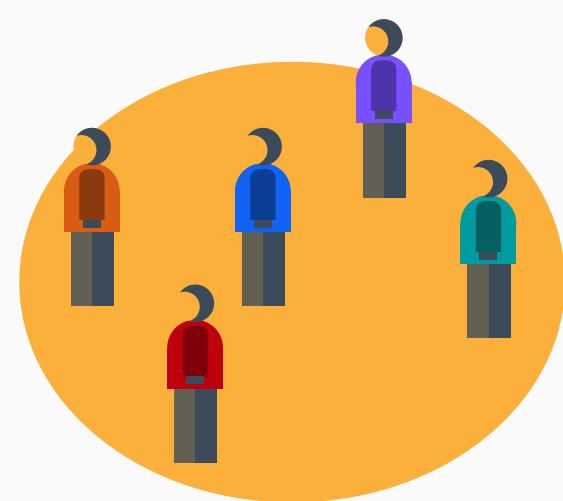


$$\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2)$$

# FREQUENTIST FRAMEWORK

One population with unknown parameters. Estimates vary across different hypothetical samples.

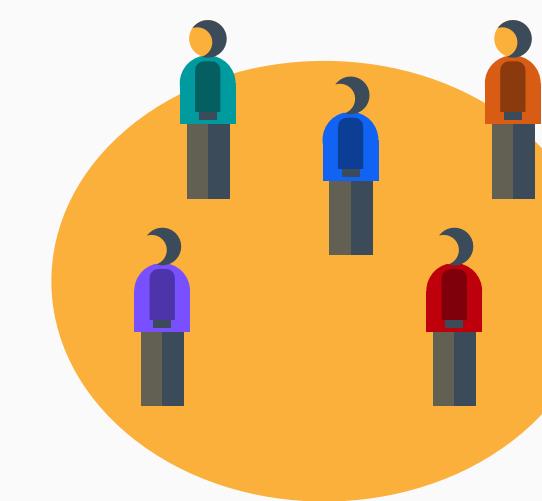
$$\beta = (\beta_0, \beta_1, \beta_2)$$



$$\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2)$$



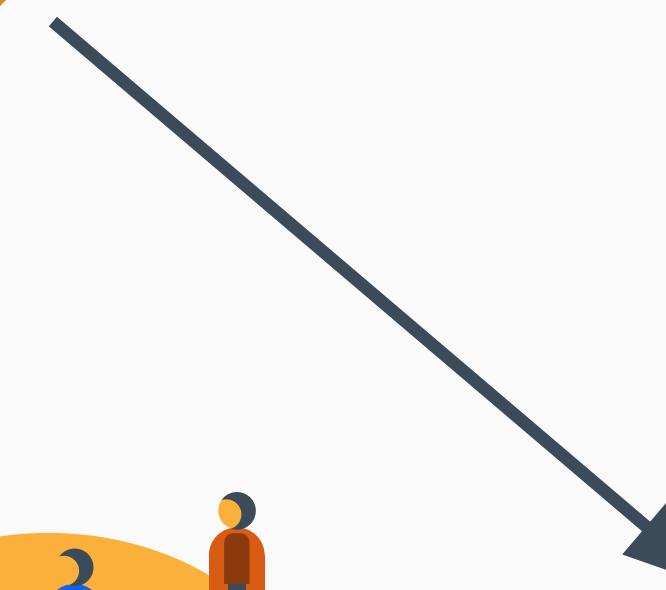
$$\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2)$$



$$\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2)$$



$$\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2)$$



# BAYESIAN STATISTICAL PARADIGM

---

- The Bayesian paradigm is defined by the idea that there is one sample of size N that we observe (no hypothetical samples)
- We imagine numerous hypothetical populations that could have produced our particular sample of data, each with unique parameter values (e.g., many values of  $\beta_0$ ,  $\beta_1$ , and  $\beta_2$ )
- The sample data are locked into place, and parameter values vary across different hypothetical populations

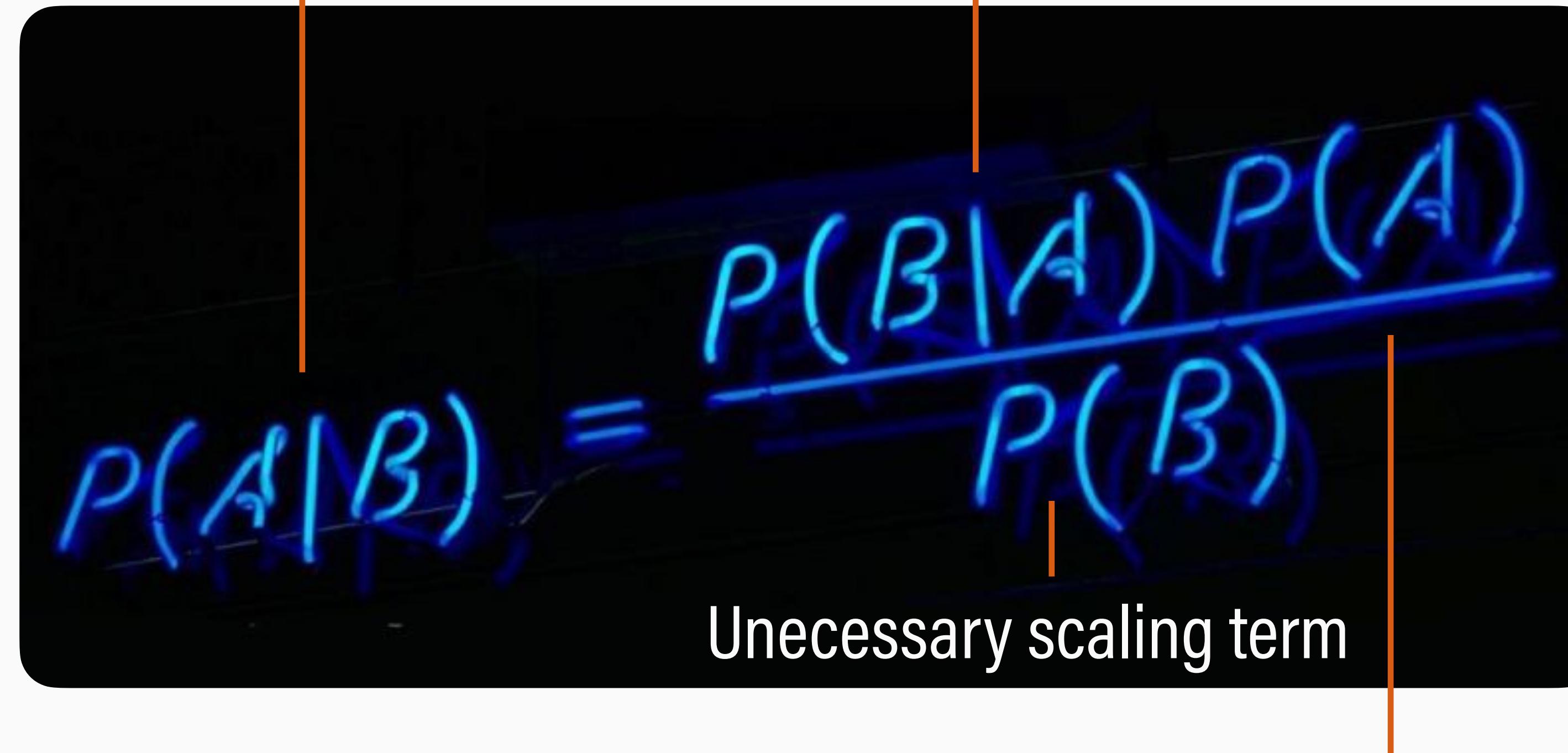
# BAYES' THEOREM

Posterior = parameters (A) given the data (B)

Frequentist likelihood = data (B) given the parameters (A)

$$P(A|B) = \frac{P(B|A) P(A)}{P(B)}$$

Unecessary scaling term

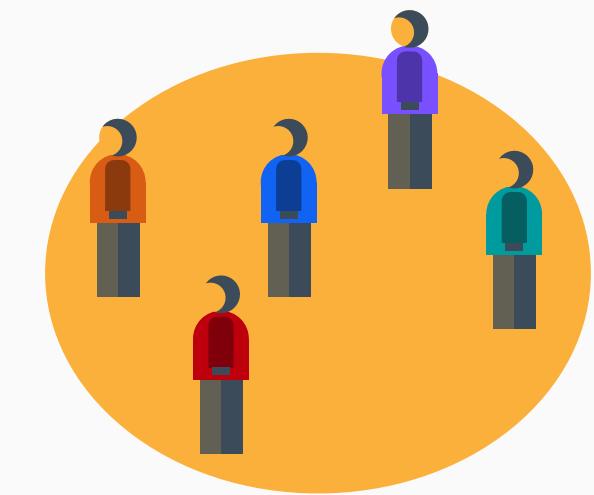


Prior = a priori belief about parameters (A)

# BAYESIAN FRAMEWORK

---

One sample of size N



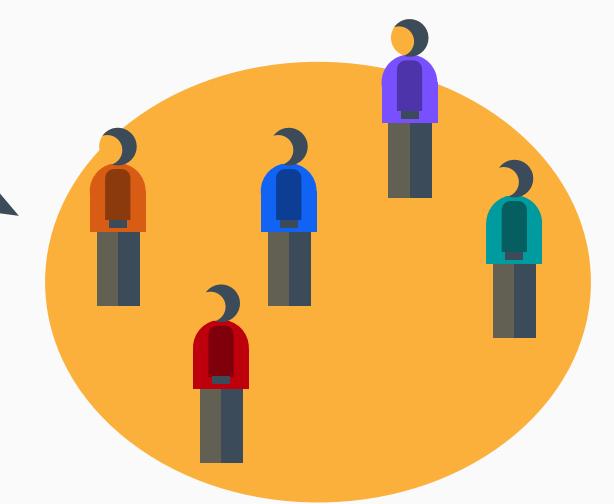
# BAYESIAN FRAMEWORK

---

$$\beta = (\beta_0, \beta_1, \beta_2)$$



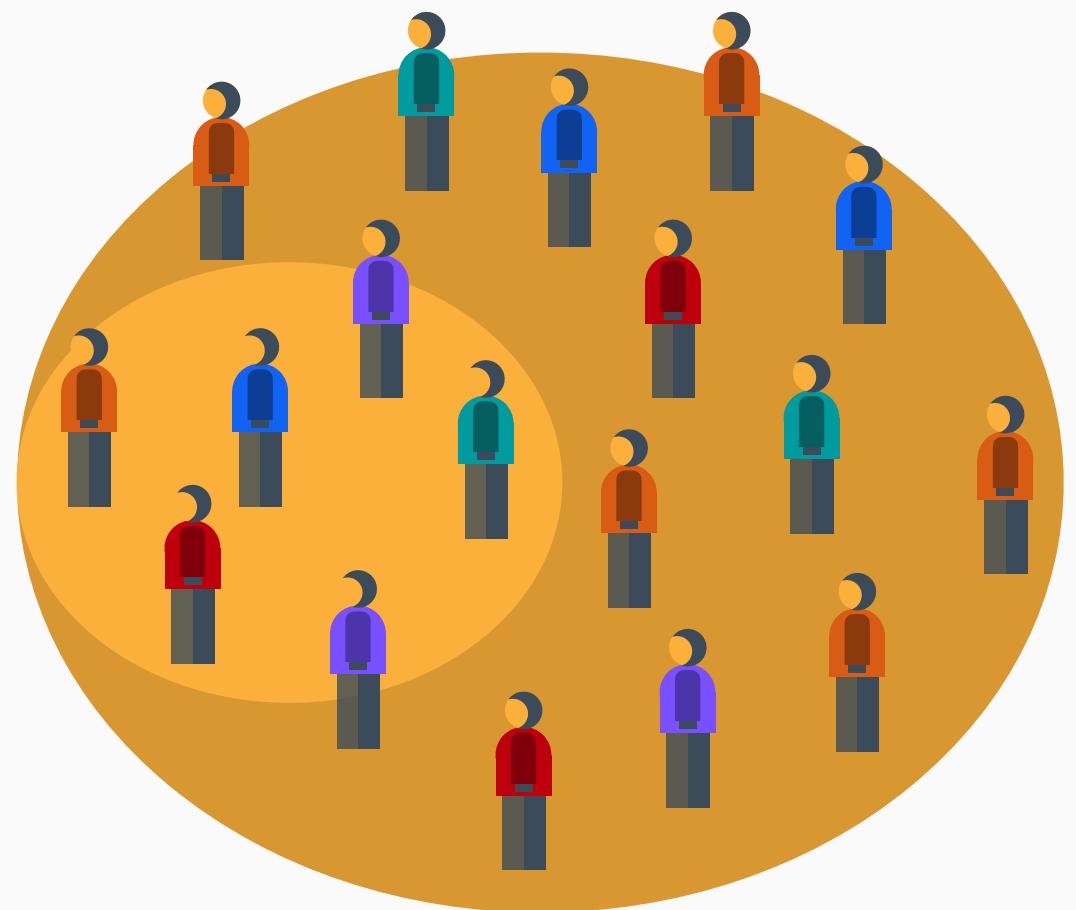
One sample of size N. Parameter  
values vary across different  
hypothetical populations.



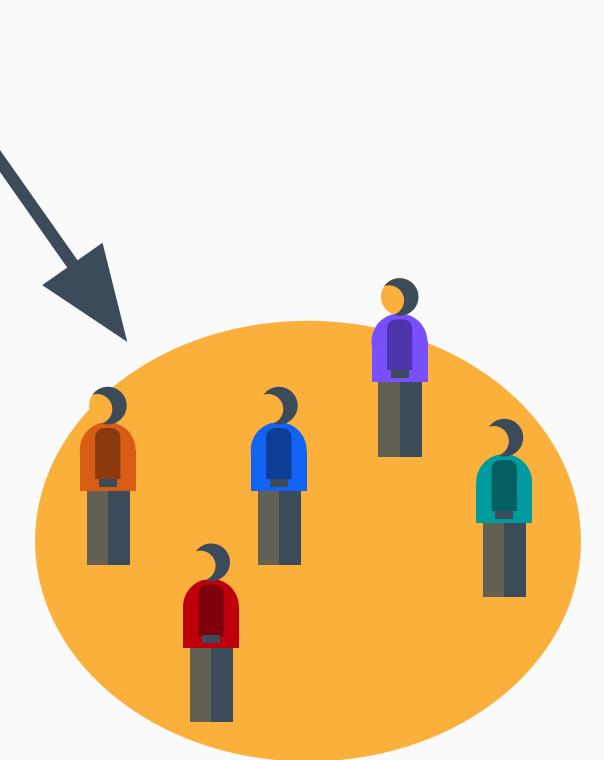
# BAYESIAN FRAMEWORK

---

$$\beta = (\beta_0, \beta_1, \beta_2)$$



$$\beta = (\beta_0, \beta_1, \beta_2)$$

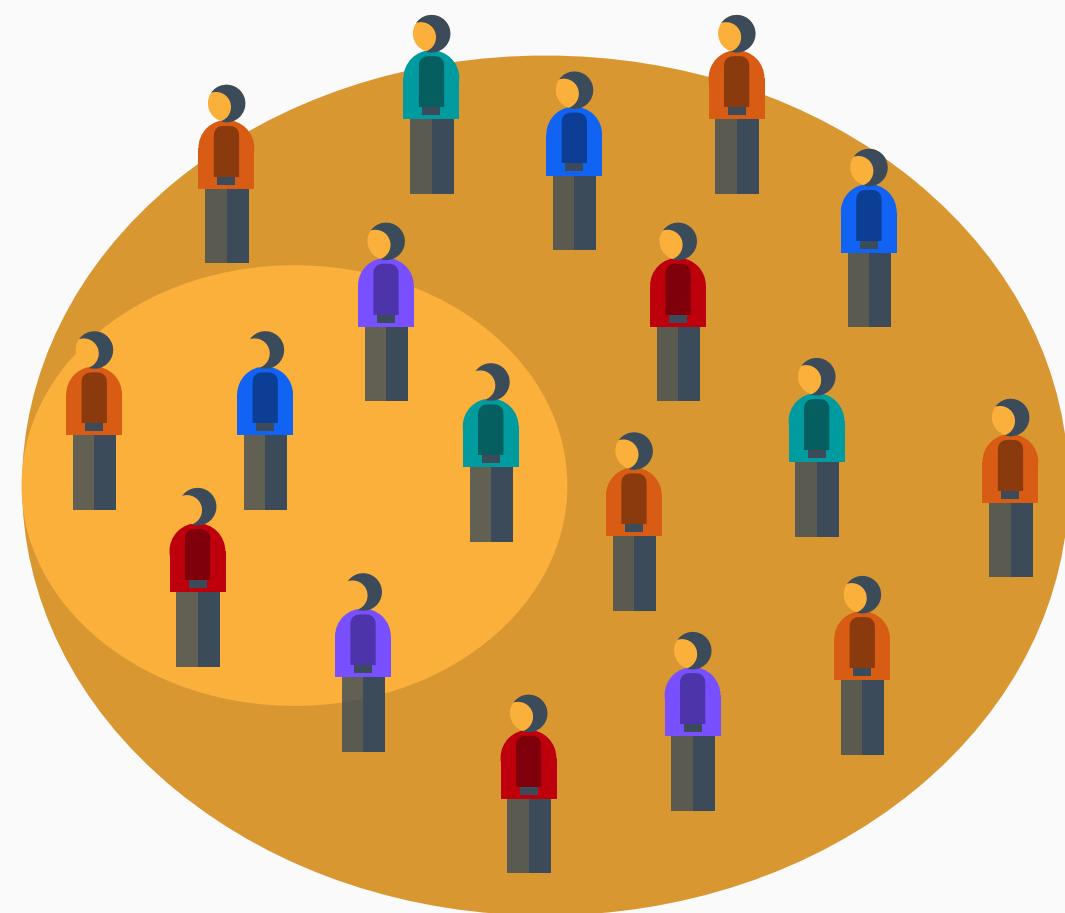


One sample of size N. Parameter  
values vary across different  
hypothetical populations.

# BAYESIAN FRAMEWORK

---

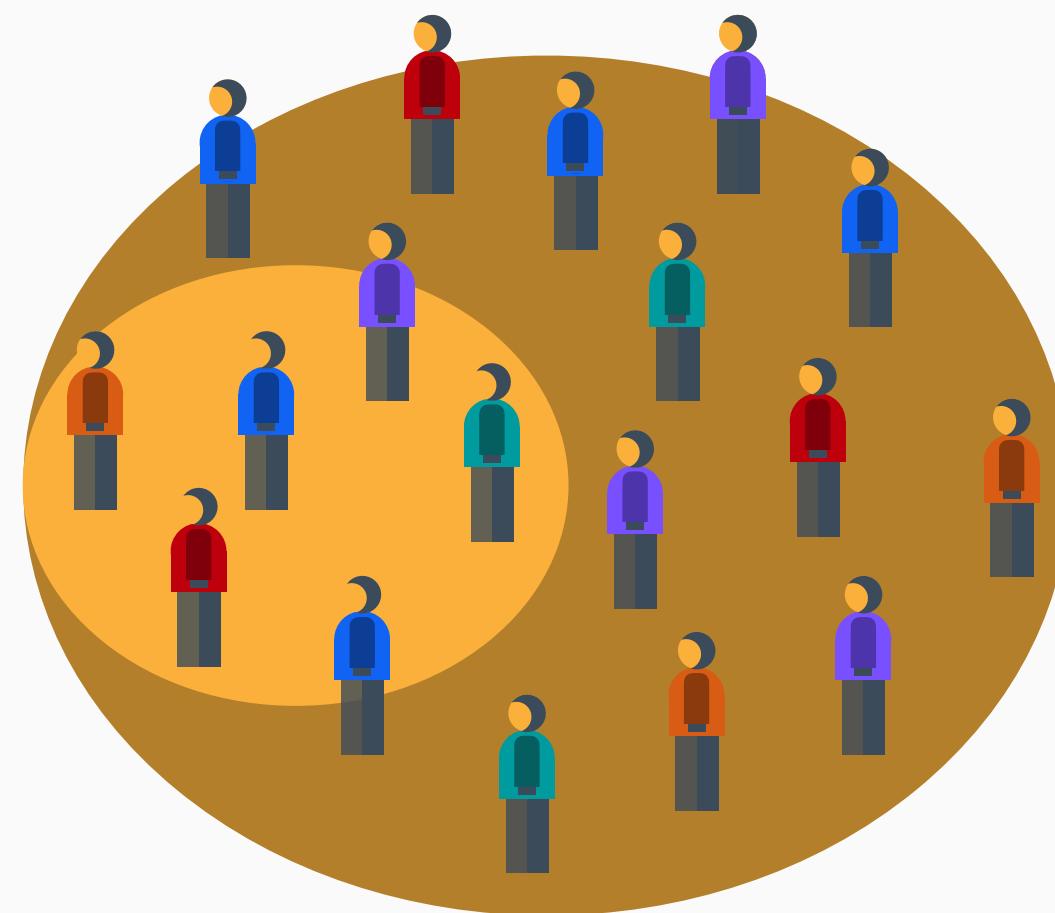
$$\beta = (\beta_0, \beta_1, \beta_2)$$



$$\beta = (\beta_0, \beta_1, \beta_2)$$



$$\beta = (\beta_0, \beta_1, \beta_2)$$

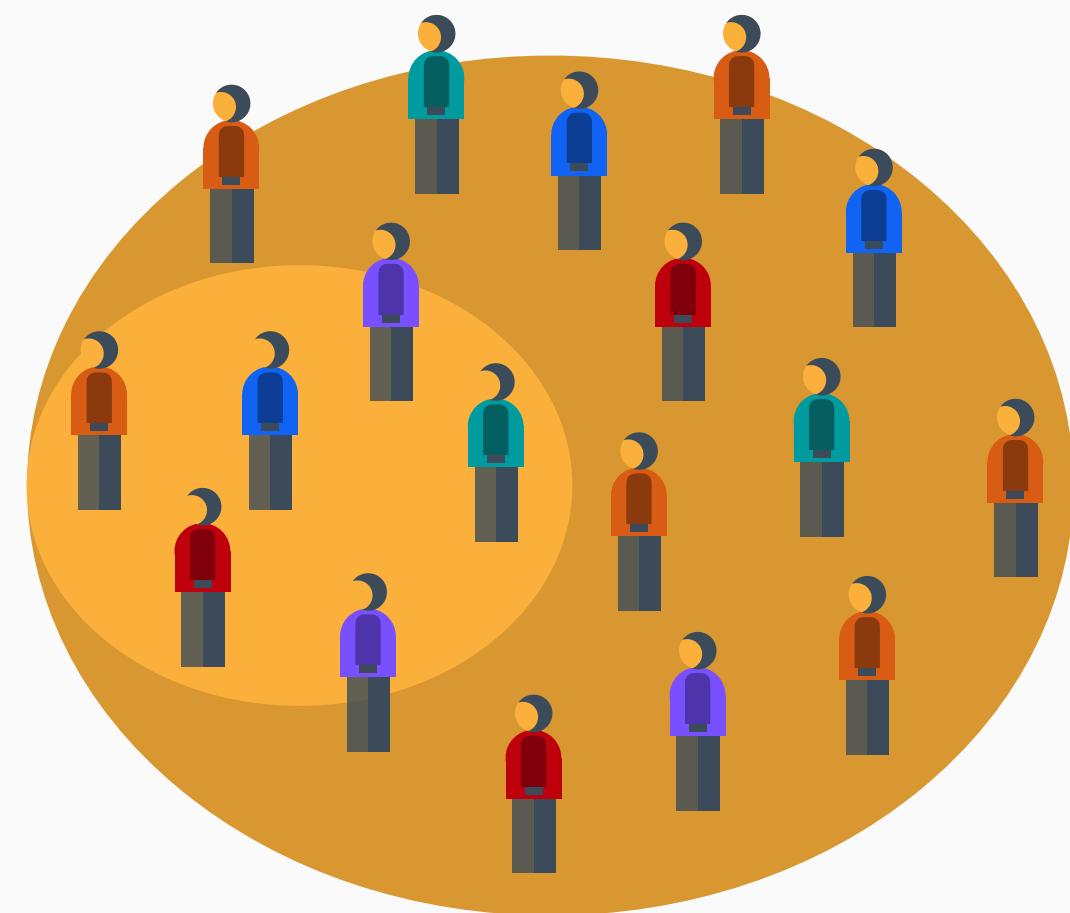


One sample of size N. Parameter values vary across different hypothetical populations.

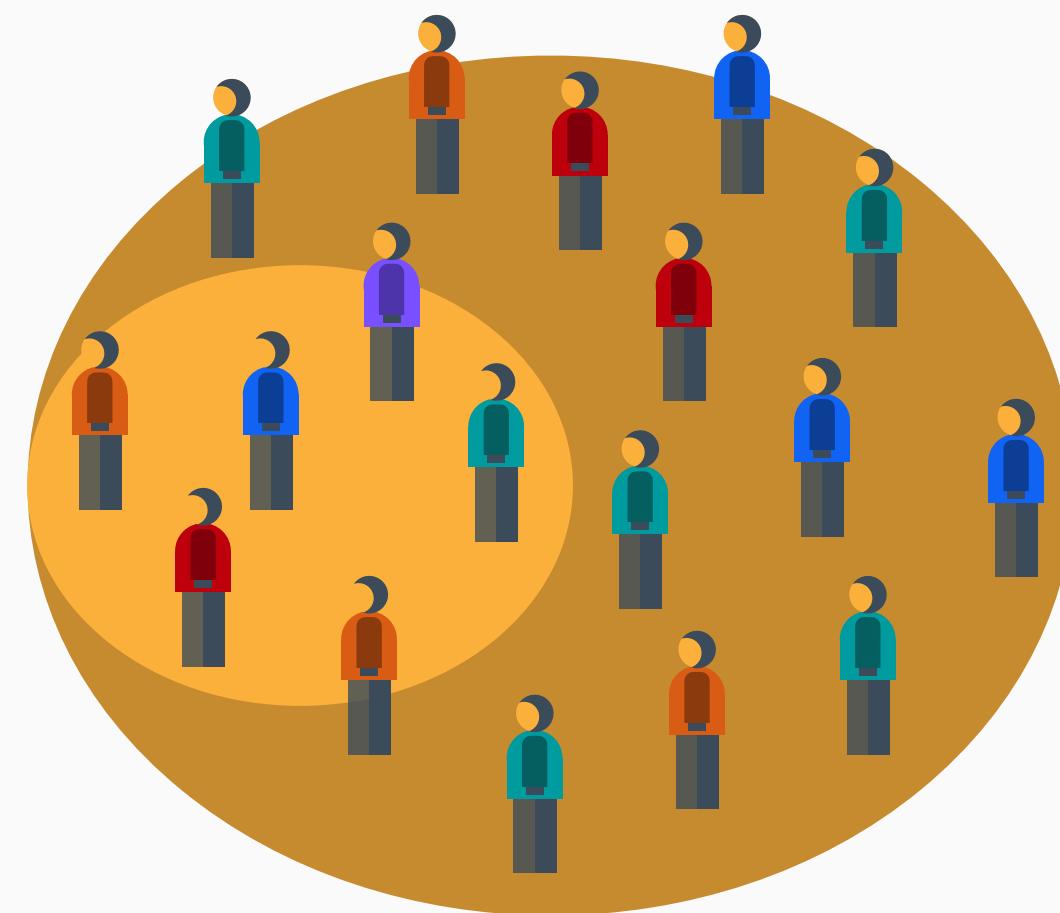
# BAYESIAN FRAMEWORK

---

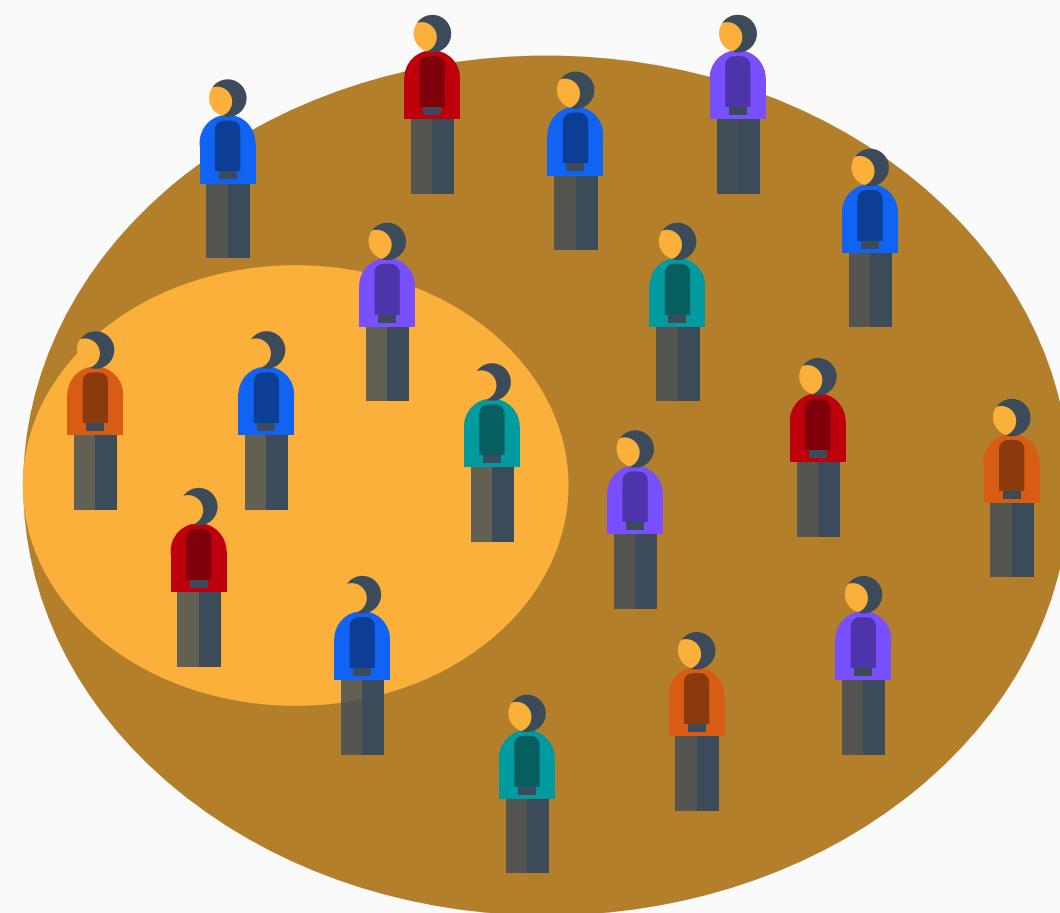
$$\beta = (\beta_0, \beta_1, \beta_2)$$



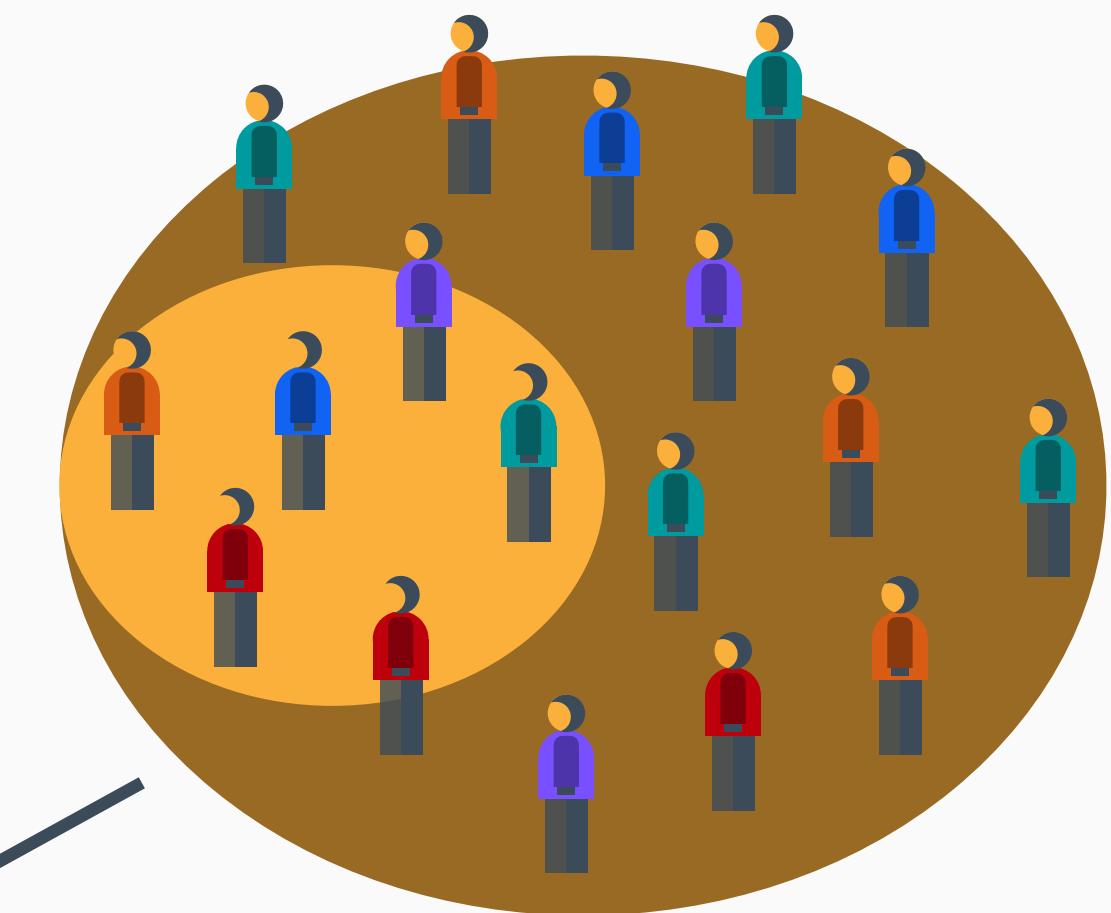
$$\beta = (\beta_0, \beta_1, \beta_2)$$



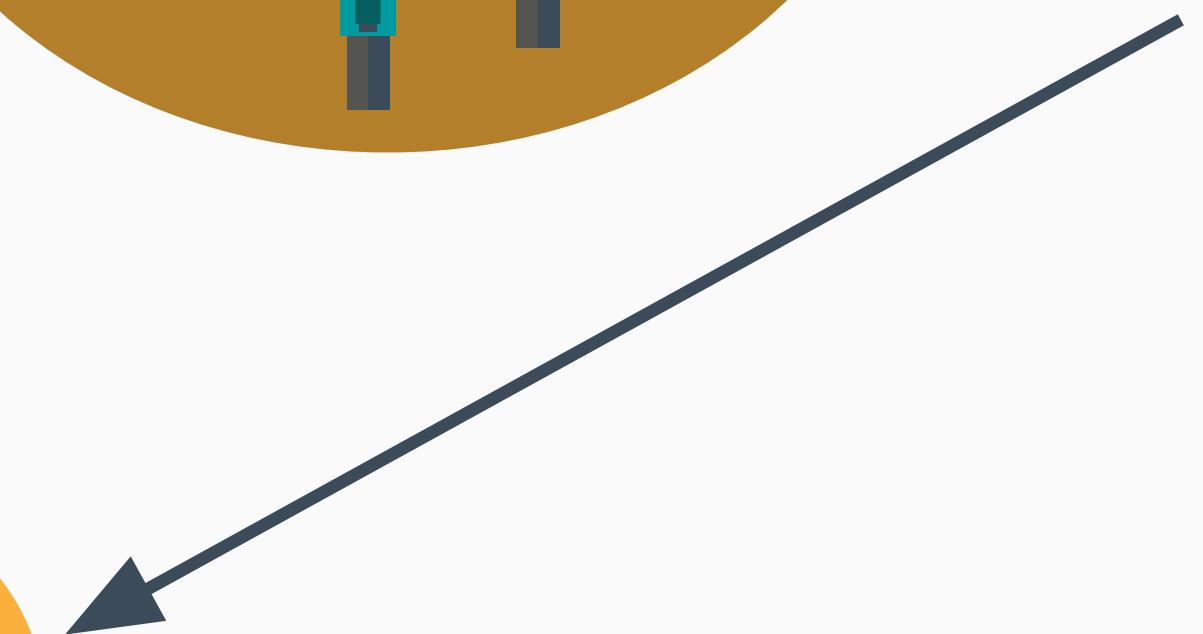
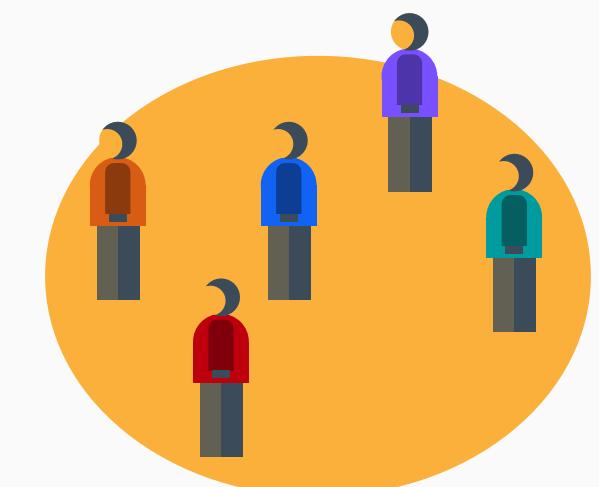
$$\beta = (\beta_0, \beta_1, \beta_2)$$



$$\beta = (\beta_0, \beta_1, \beta_2)$$



One sample of size N. Parameter values vary across different hypothetical populations.



# FREQUENTIST VS. BAYESIAN SUMMARY

---

## Frequentist

- The parameter is a fixed quantity, estimates vary across different samples
- Statements about probability, precision, and confidence refer to estimates
- Probability = the long run frequency of an event across many different samples

## Bayesian

- There is a single sample, and parameters vary across different populations
- Statements about probability, precision, and intervals refer to the parameter
- Probability = our degree of certainty about a parameter after analyzing data

# OUTLINE

- 1 Analysis Example
- 2 Frequentist vs. Bayesian Statistical Paradigms
- 3 Maximum Likelihood Estimation
- 4 MCMC Estimation
- 5 MCMC Diagnostics
- 6 Checking MLM Assumptions

# MAXIMUM LIKELIHOOD ESTIMATION

---

- Maximum likelihood (ML) estimation identifies the population parameters that are most likely to have produced the data
- Like ordinary least squares regression, maximum likelihood estimates (MLEs) minimize the sum of squared residuals
- Normal curve functions define two layers of MLM residuals

# ITERATIVE OPTIMIZATION ALGORITHMS

---

- Unlike OLS, multilevel models do not have closed-form equations where you plug in data and solve for the estimates
- ML requires iterative optimizers (Newton's or EM algorithm) that cyclically update the parameters
- Each successive step adjust estimates to improve fit (decrease residuals) until the data-model fit is maximized

# REVIEW: ESTIMATING VARIANCES

---

- OLS can be viewed as two-steps: (1) Use all  $N$  data points to estimate coefficients, (2) subtract the “spent” degrees of freedom (df) from the  $N$  and estimate the residual variance
- Computing the variance with  $N - 1$  as a divisor gives unbiased estimates, whereas using  $N$  attenuates variability estimates
- The degrees of freedom adjustment for residual variances in regression similarly uses  $N - \# \text{ predictors} - 1$

# FIML VERSUS REML ESTIMATORS

---

- Full information maximum likelihood (FIML) does use bias-reducing degrees of freedom adjustments
- Level-2 variance estimates are attenuated when the number of clusters is small because no degrees-of-freedom adjustment is applied (e.g., they use  $J$  rather than  $J - 1$ )
- Restricted maximum likelihood (REML) addresses this bias, estimating variances with df-like adjustments

# FIML VS. REML COMPARISON

---

- FIML estimates of the level-2 residual variance are slightly lower because they do not adjust for degrees of freedom spent estimating the coefficients

Parameter	REML		FIML		Bayesian MCMC	
	Est.	Std. Err.	Est.	Std. Err.	Est.	Std. Err.
Fixed intercept	5.05	0.12	5.05	0.12	5.00	0.13
Sleep (within-person)	0.17	0.01	0.17	0.01	0.17	0.01
Sleep (between-person)	0.58	0.09	0.58	0.08	0.60	0.09
Random intercept variance	<b>1.83</b>	--	<b>1.80</b>	--	1.84	0.25
Residual within-person variance	1.31	--	1.30	--	1.31	0.04

# OUTLINE

- 1 Analysis Example
- 2 Frequentist vs. Bayesian Statistical Paradigms
- 3 Maximum Likelihood Estimation
- 4 MCMC Estimation
- 5 MCMC Diagnostics
- 6 Checking MLM Assumptions

# WHY CHOOSE MCMC?

---

- MCMC readily handles complex missing data problems, including:
  - Mixed metrics (normal, ordinal, nominal, skewed, count, latent)
  - Nonlinear effects (interactions, curvilinear effects)
  - Multilevel data (random coefficients, interactions)
  - Latent variable modeling (interactions)
- FIML estimators for these scenarios are far more limited

# MARKOV CHAIN MONTE CARLO (MCMC)

---

- MCMC estimation samples model parameters and the missing values from distributions of plausible values
- MCMC breaks a complex problem involving multiple unknowns (parameters and missing values) into separate steps
- Each step estimates one unknown at a time, treating the current values of all other quantities as known constants

# MCMC ALGORITHM

---

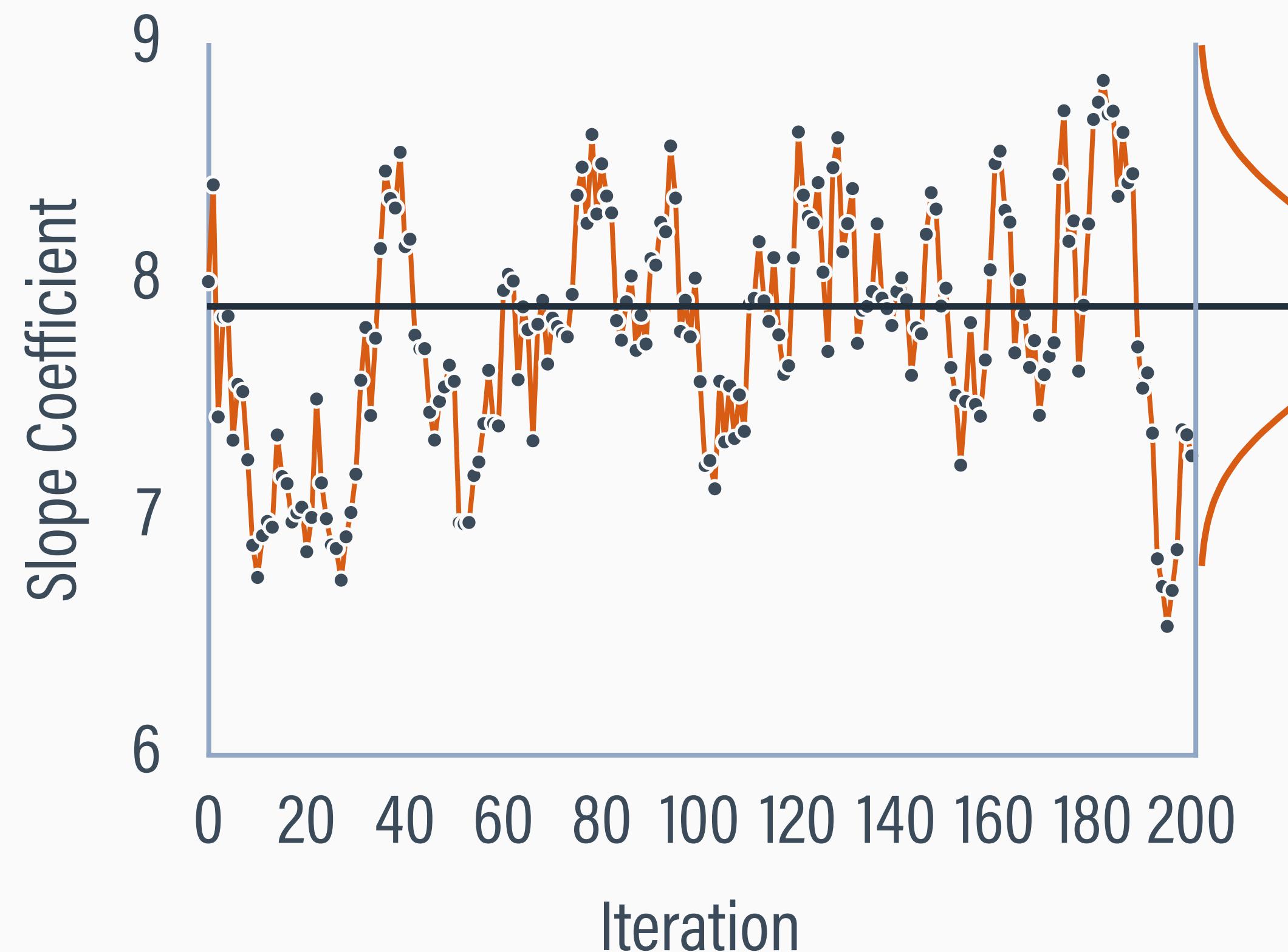


- » Do for  $t = 1$  to  $T$  iterations
  - » Estimate regression coefficients
  - » Estimate random intercept residuals
  - » Estimate level-2 intercept variance
  - » Estimate level-1 residual variance
  - » Impute missing data
- » Repeat

# MEANING OF ESTIMATION

---

- MCMC uses computer simulation to “sample” parameters from a distribution
- Estimates continually vary across iterations in a random pattern
- Each iteration gives plausible parameter values that could have produced our data



# DISAGGREGATED ANALYSIS

---

- Affect regressed on within- and between-person sleep

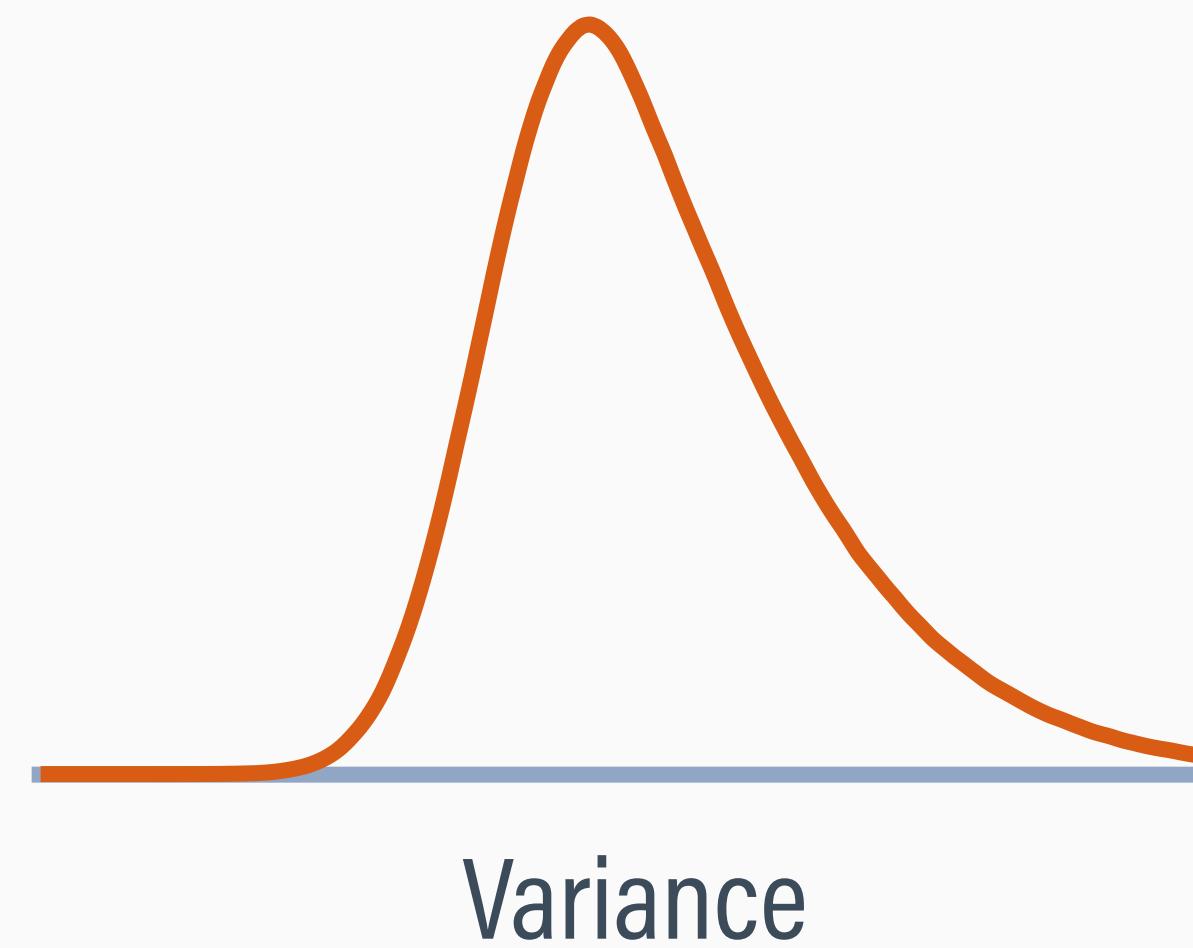
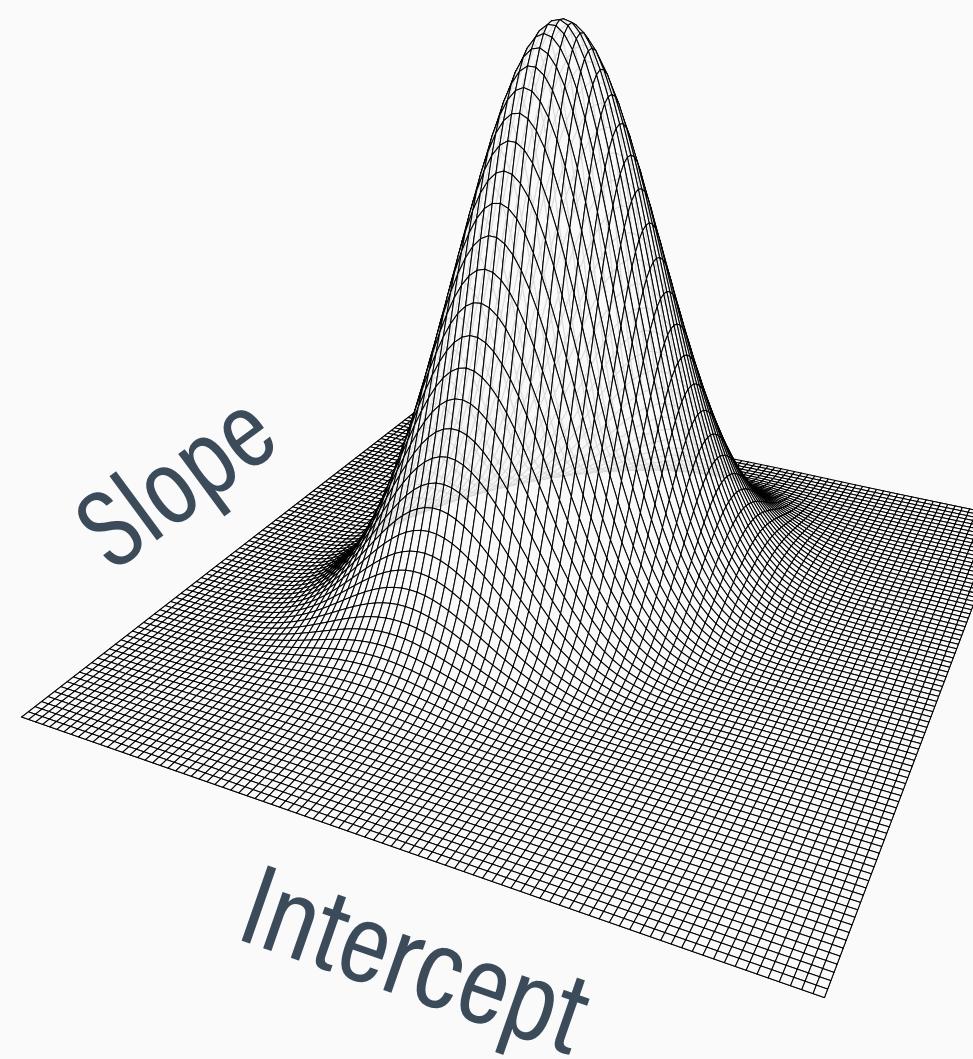
$$p\text{affect}_{ij} = \beta_0 + \beta_1(\text{sleep}_{ij}^W) + \beta_2(\text{sleep}_j^B) + u_{0j} + \varepsilon_{ij}$$

- Each iteration yields population parameter values that could have plausibly produced this same of data
- The goal is to summarize the parameter distributions

# PARAMETER-GENERATING DISTRIBUTIONS

---

- MCMC draws coefficients from a multivariate normal distribution, with least-squares estimates defining shape
- MCMC draws variances from an inverse gamma distribution with its shape determined by the df and residual SS



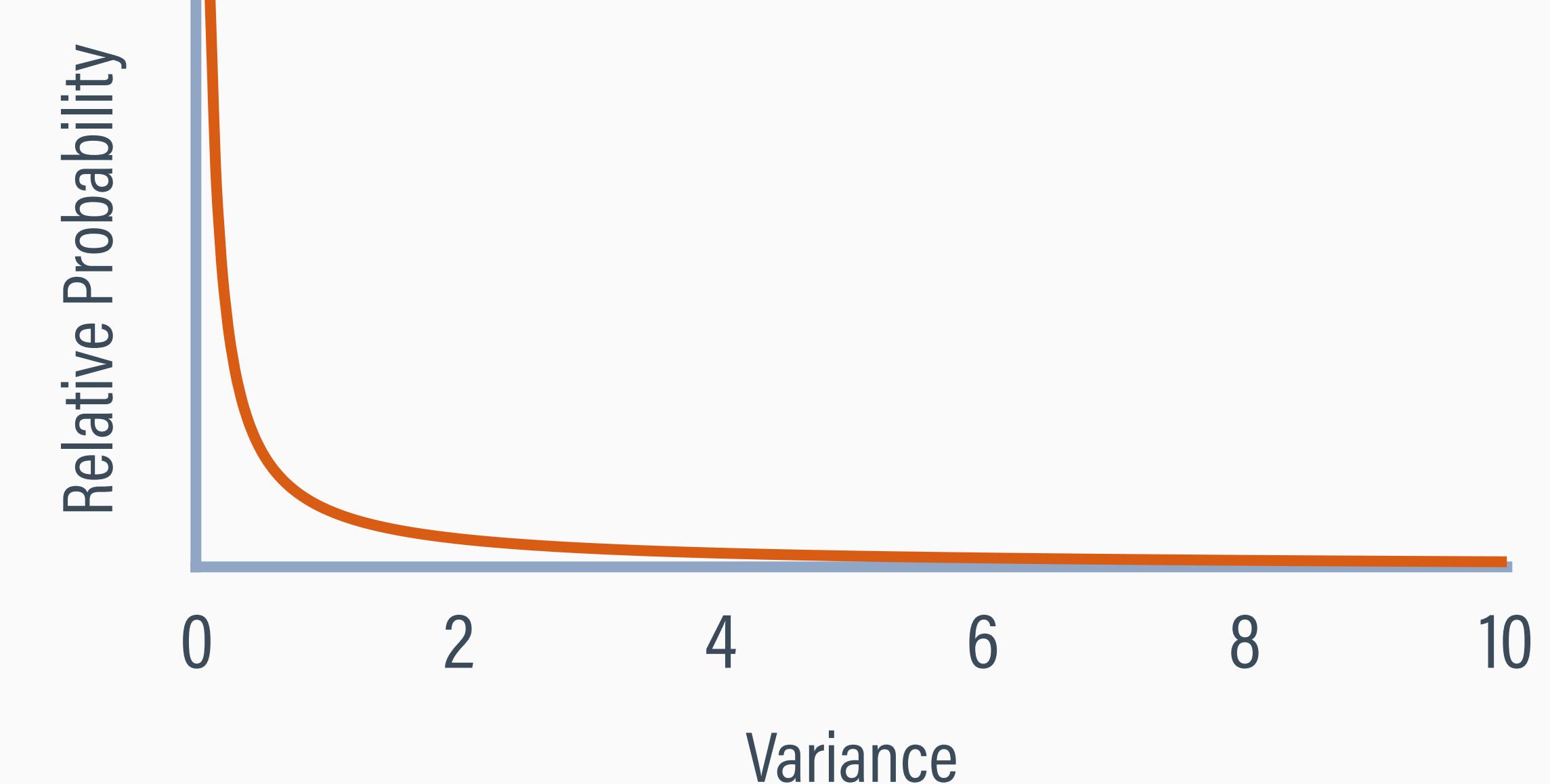
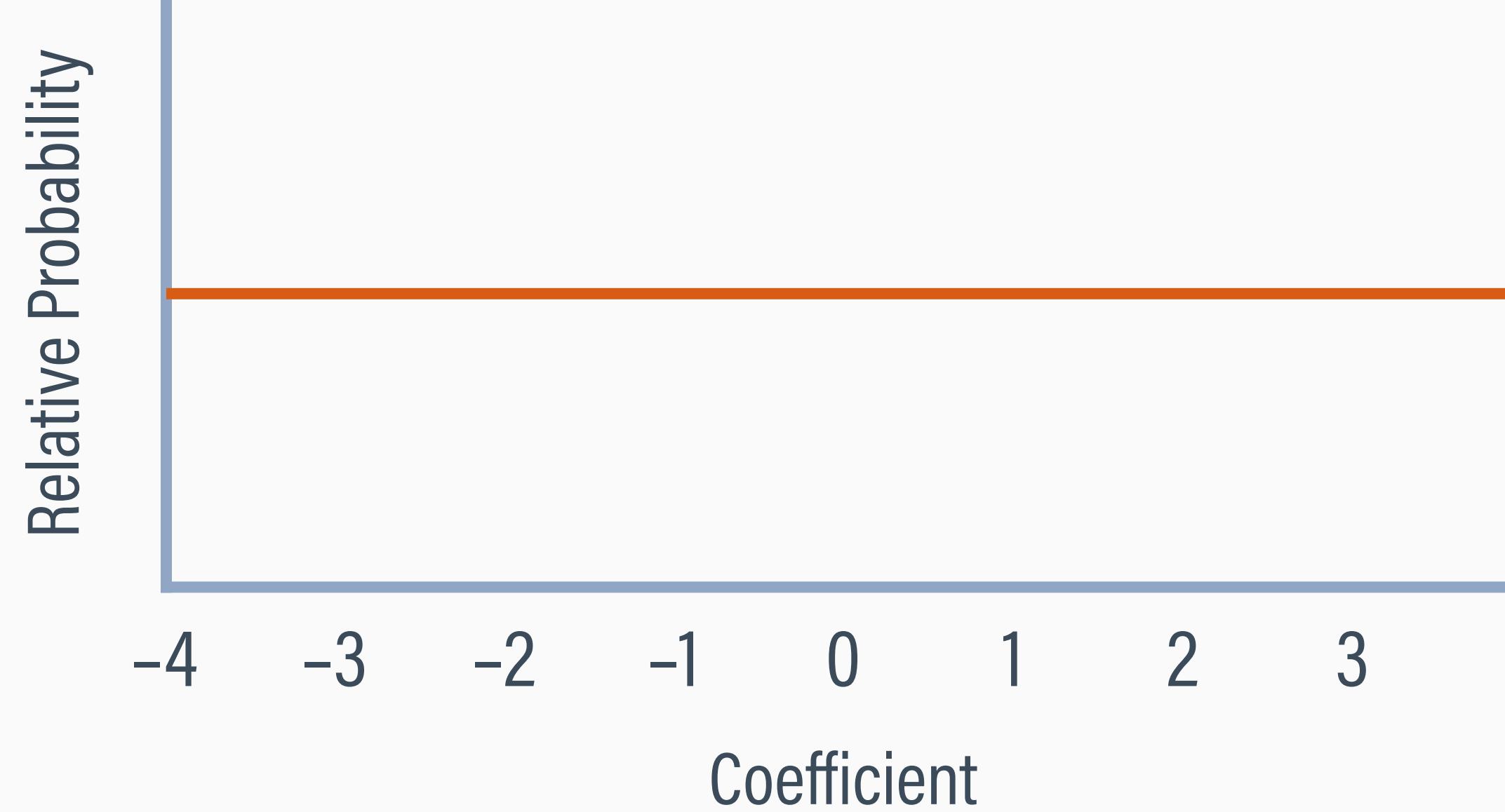
# PRIOR DISTRIBUTIONS

---

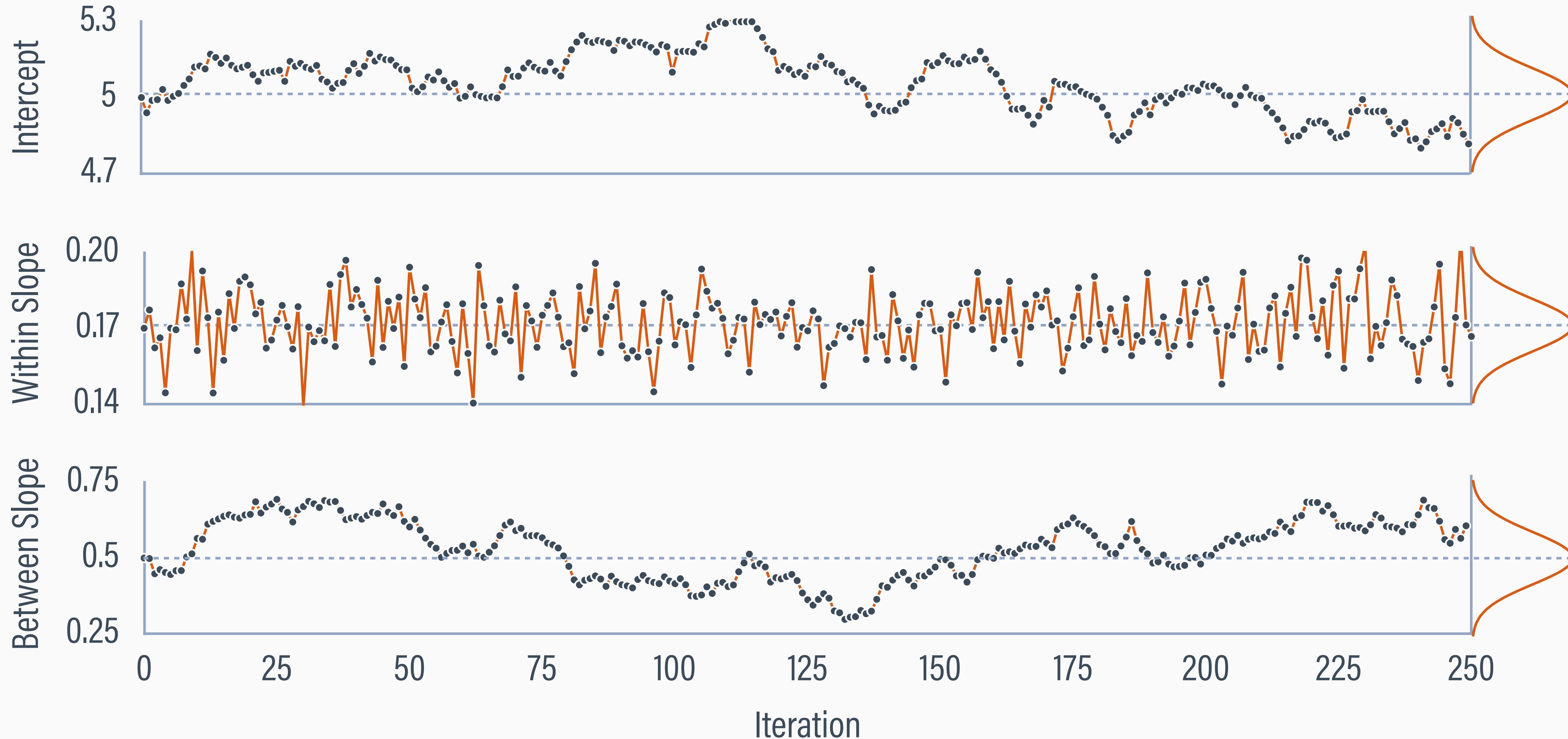
- Bayesian analyses require prior distributions that encode our beliefs about the parameter values prior to analyzing the data
- Conceptually, prior distributions function like secondary inputs that augment the data during estimation
- It is common to use non-informative (diffuse) priors that impart as little information as possible (let the data do the talking)

# PRIOR DISTRIBUTIONS

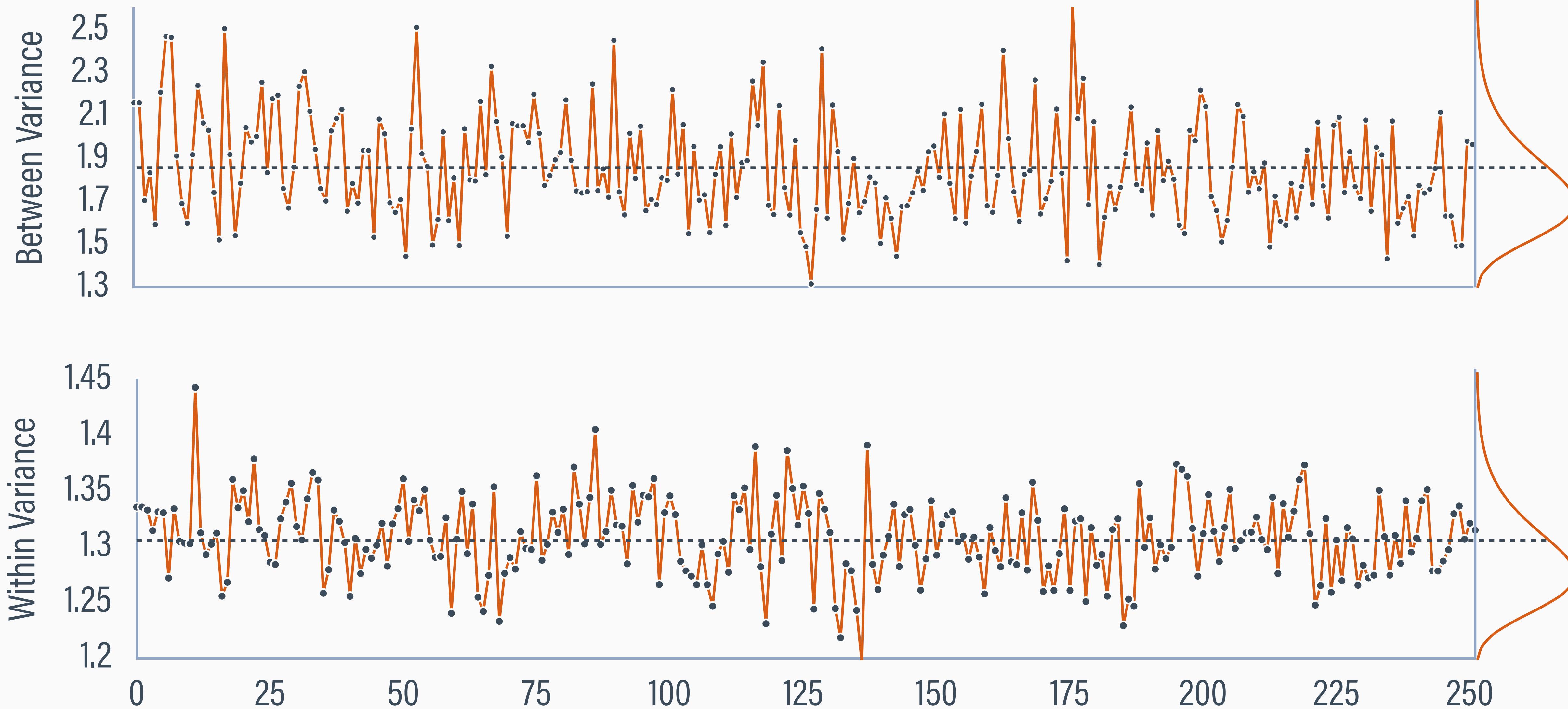
- A diffuse prior for means and coefficients conveys that all possible parameter values are equally likely a priori
- Diffuse priors for variances are slightly informative, and different options function like df adjustments in regression



# COEFFICIENTS FROM 250 MCMC CYCLES



# VARIANCES FROM 250 MCMC CYCLES



# SUMMARIZING MCMC ESTIMATES

---

- MCMC iterates for thousands of cycles, and each cycle produces estimates based on one fixed-in data set
- MCMC estimation yields a distribution of parameters—called a posterior—that averages over thousands of imputations
- The posterior is a distribution of plausible parameter values that could have produced our particular data

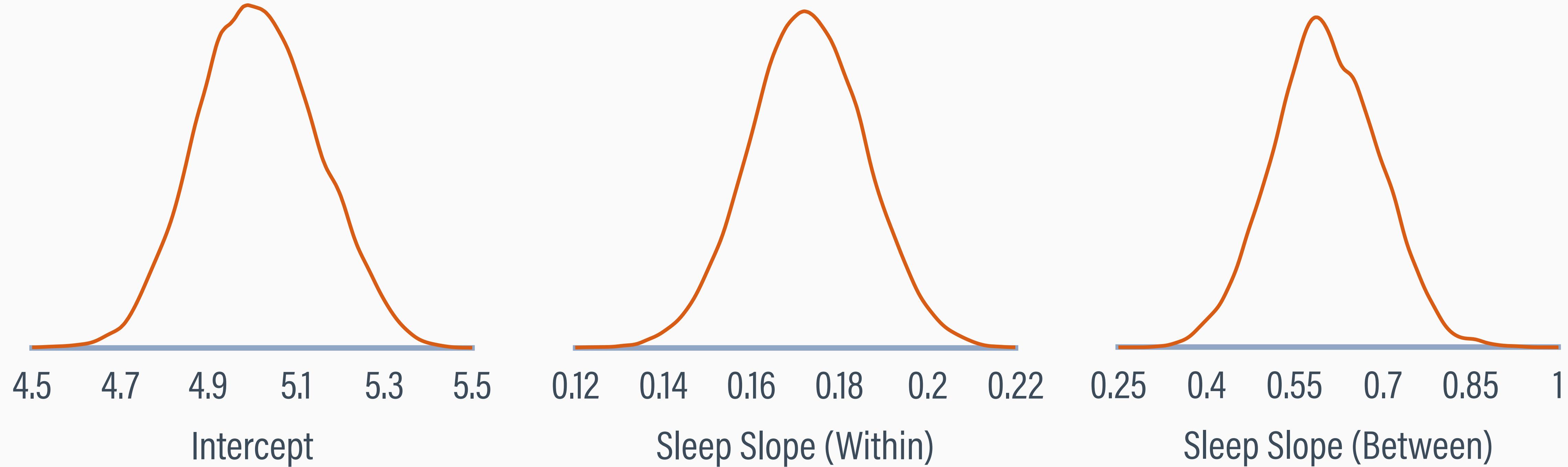
# COEFFICIENT DISTRIBUTIONS

---

Median = 5.01  
Std. Dev. = 0.14  
95% CI = (4.76, 5.28)

Median = 0.17  
Std. Dev. = 0.01  
95% CI = (0.15, 0.20)

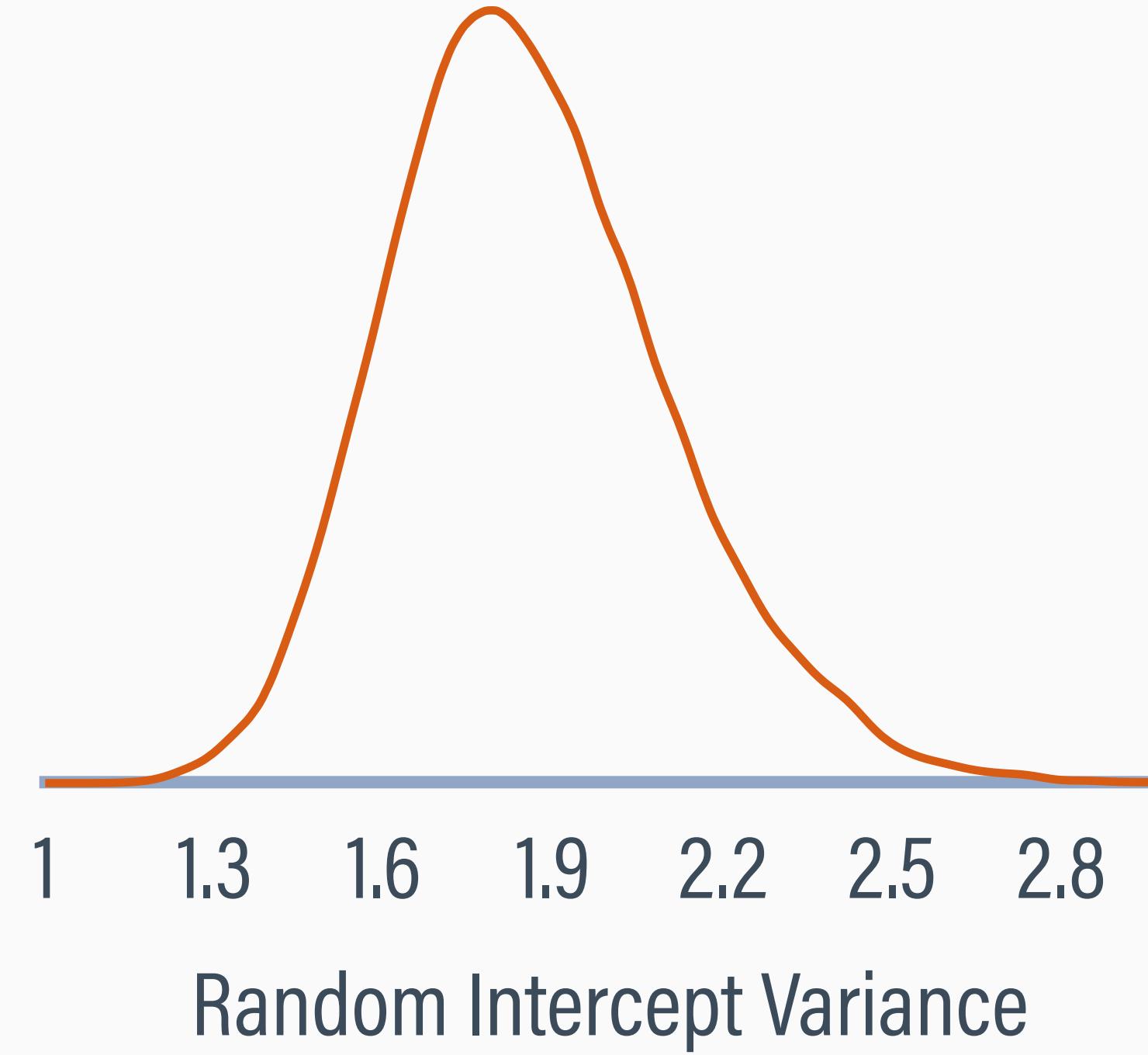
Median = 0.60  
Std. Dev. = 0.09  
95% CI = (0.43, 0.78)



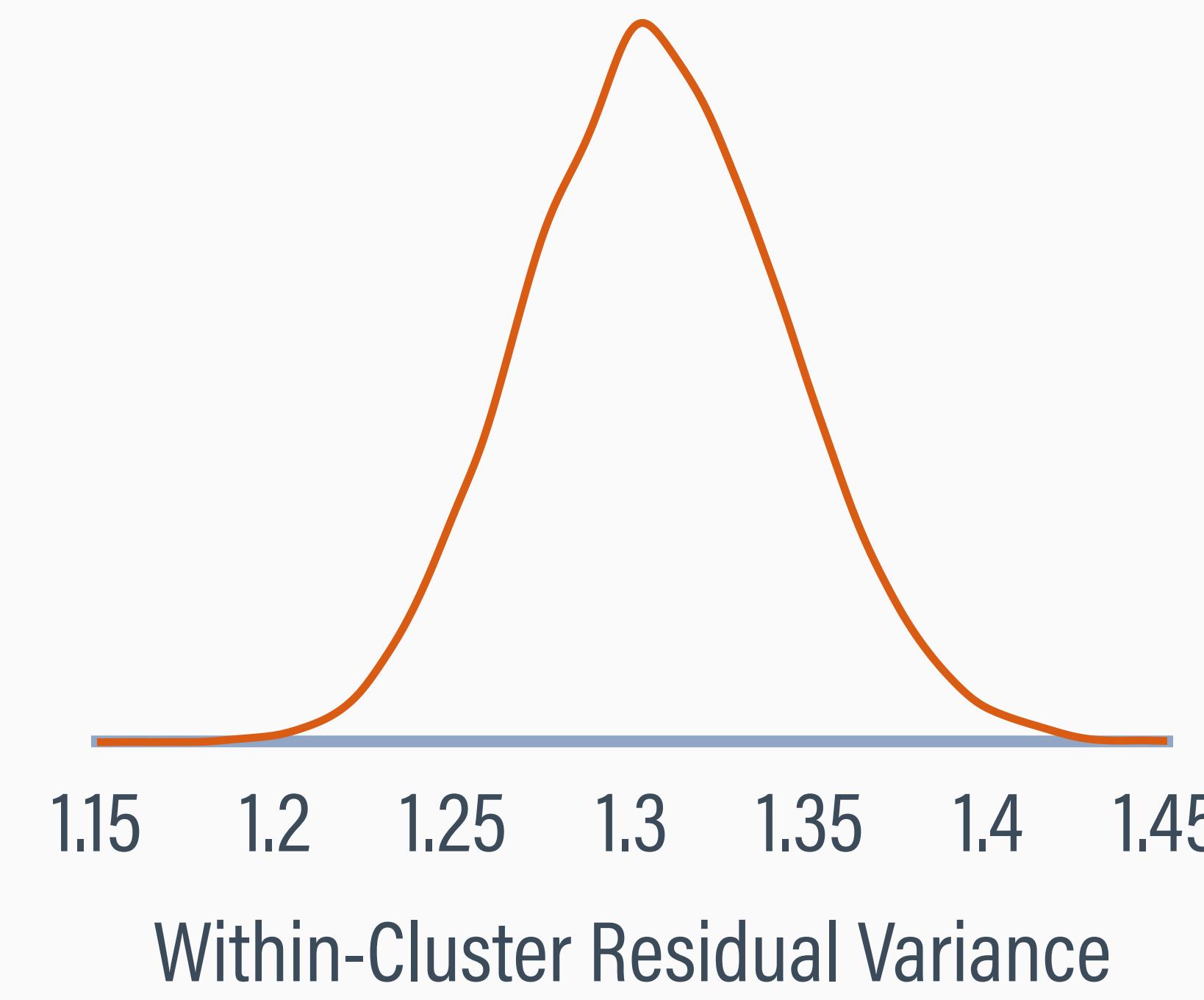
# VARIANCE DISTRIBUTIONS

---

Median = 1.84  
Std. Dev. = 0.25  
95% CI = (1.43, 2.40)



Median = 1.31  
Std. Dev. = 0.04  
95% CI = (1.24, 1.38)

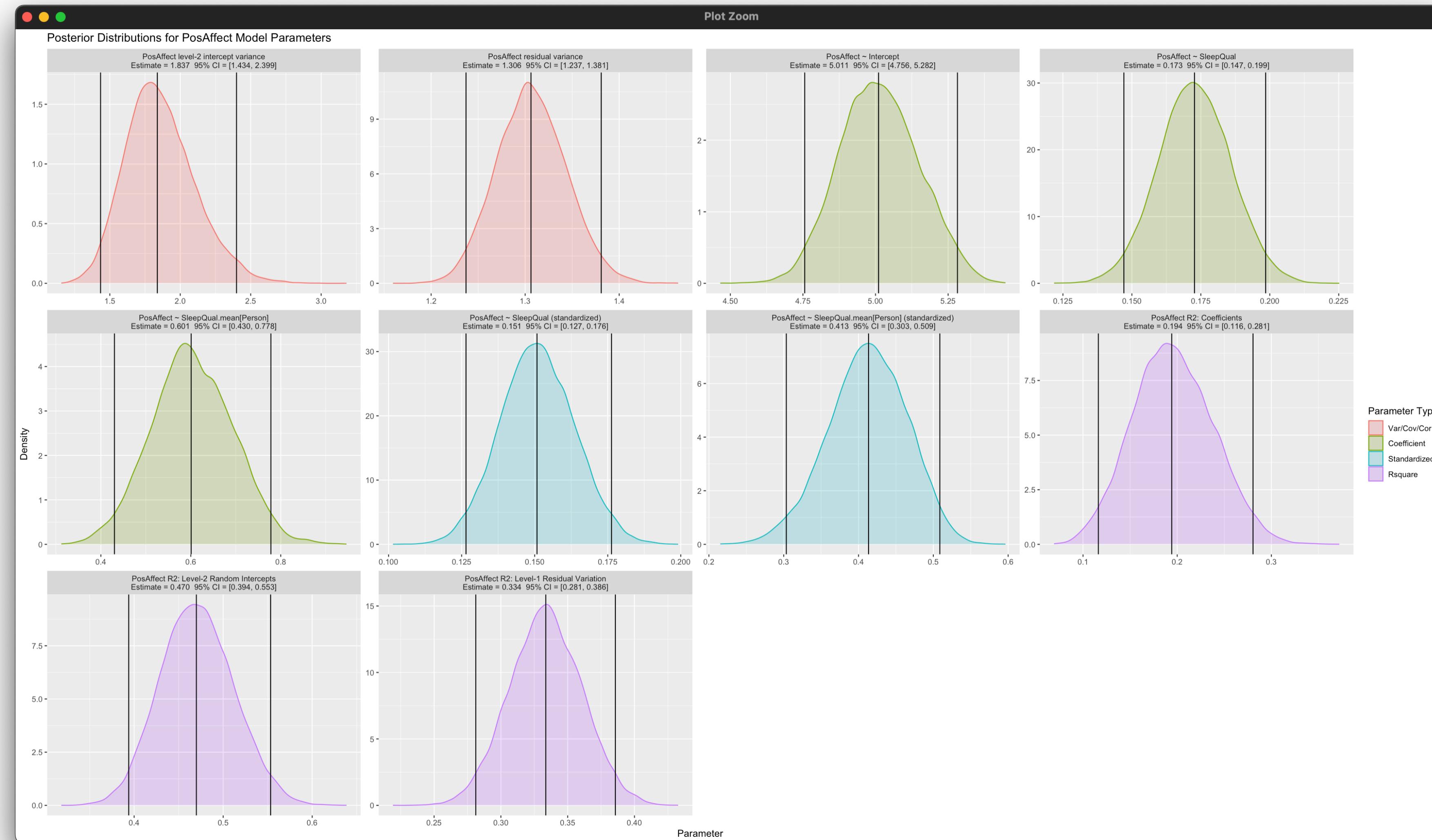


# RBLIMP SCRIPT

---

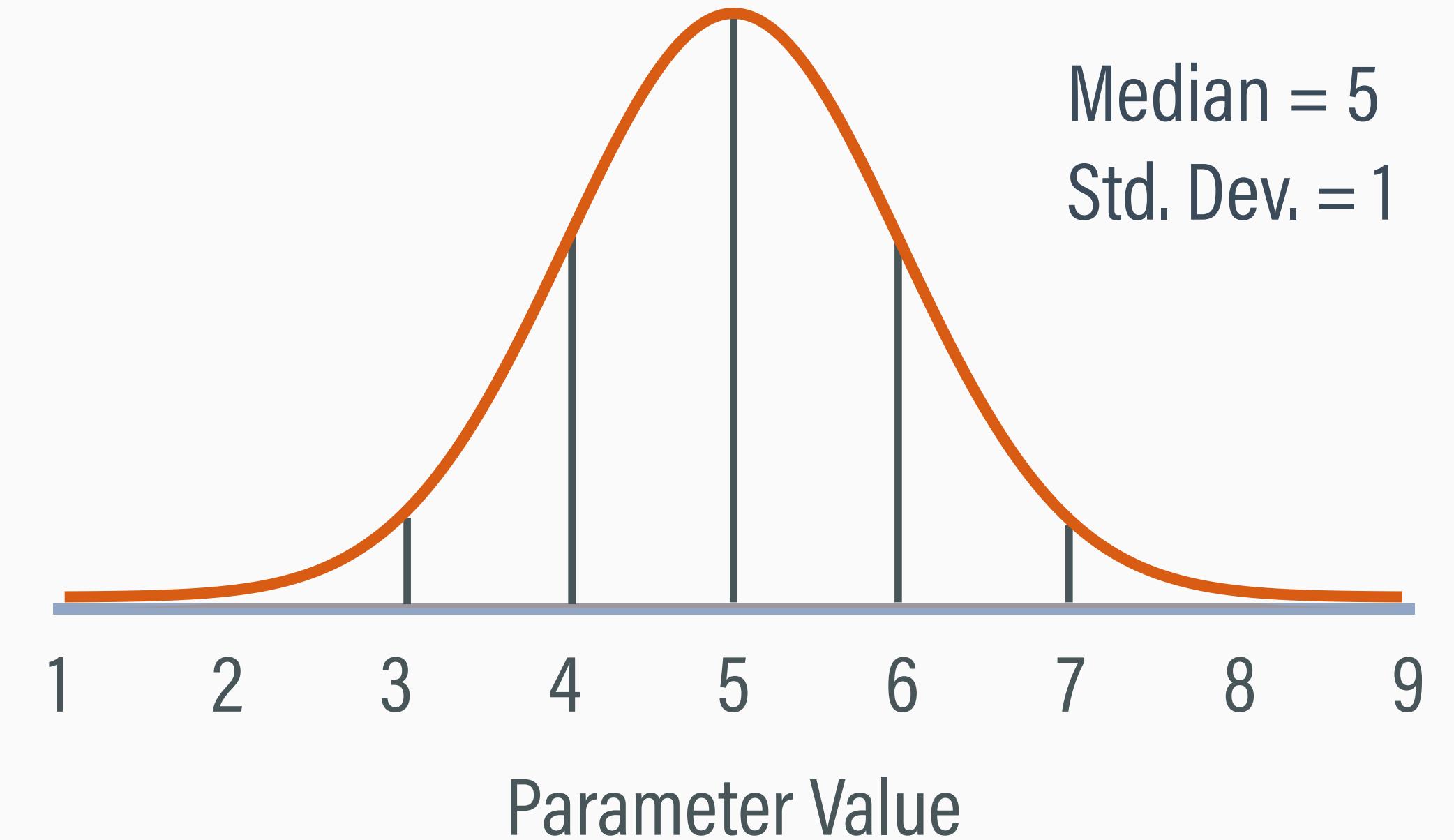
```
model <- rblimp(  
  data = PainDiary,  
  clusterid = 'Person',  
  center = 'grandmean = SleepQual.mean; groupmean = SleepQual',  
  model = 'PosAffect ~ intercept SleepQual SleepQual.mean | intercept',  
  seed = 90291,  
  burn = 10000,  
  iter = 10000)  
  
output(model)  
posterior_plot(model)
```

# PARAMETER PLOTS (RBLIMP ONLY)



# POSTERIOR MEDIAN AND STD. DEV.

- The posterior median and standard deviation quantify the most likely parameter value and uncertainty
- Analogous to a point estimate and standard error but no reference to other hypothetical samples



# BLIMP OUTPUT

---

Outcome Variable: PosAffect

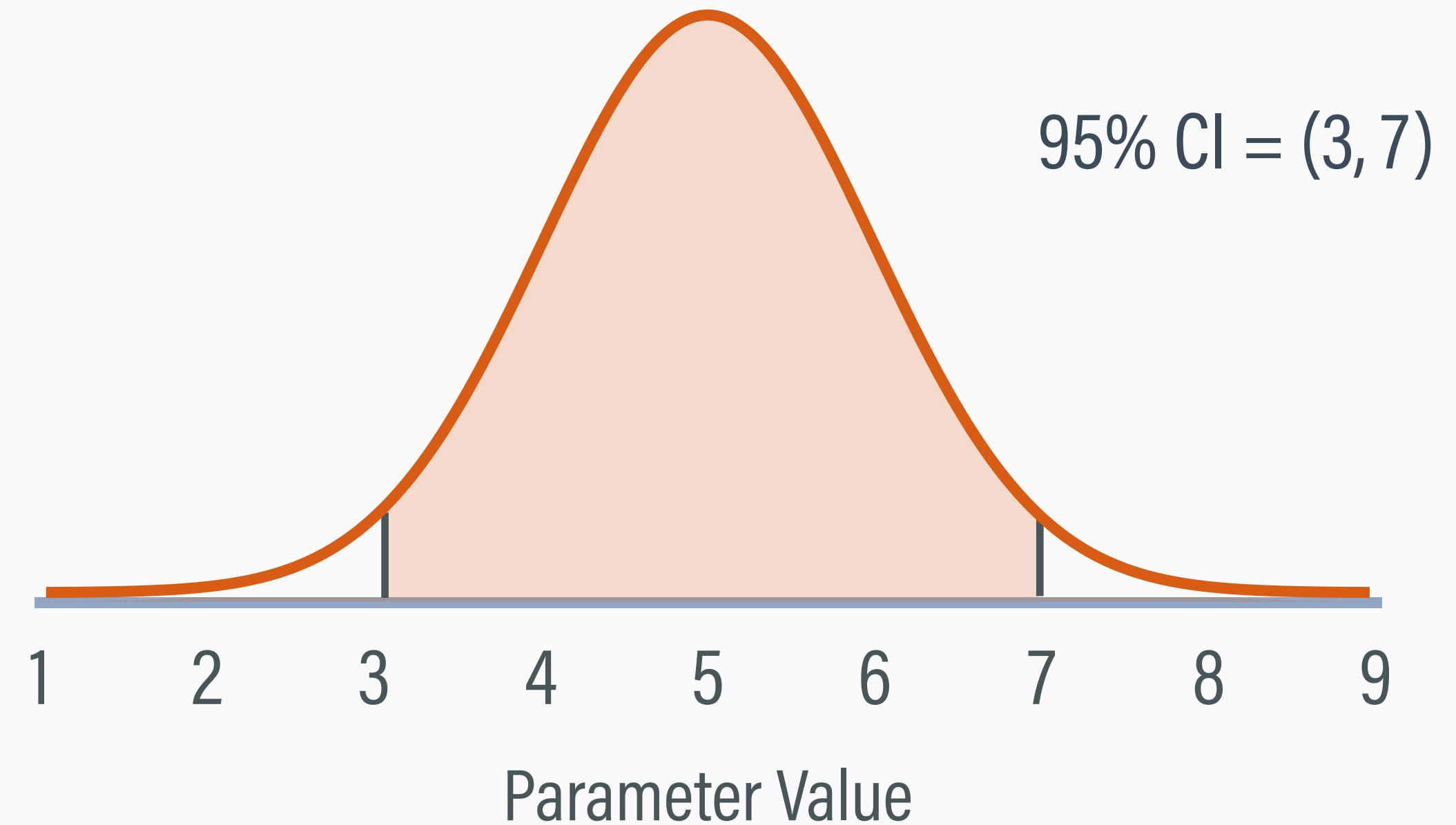
Group Mean Centered: SleepQual

Parameters	Estimate	StdDev	2.5%	97.5%	ChiSq	PValue	N_Eff
<hr/>							
Variances:							
L2 : Var(Intercept)	1.839	0.246	1.436	2.399	---	---	5051.168
Residual Var.	1.307	0.037	1.238	1.380	---	---	8227.580
<hr/>							
Coefficients:							
Intercept	4.997	0.133	4.761	5.277	1423.604	0.000	154.295
SleepQual	0.173	0.013	0.147	0.199	174.595	0.000	9302.345
SleepQual.mean[Person]	0.598	0.094	0.416	0.788	40.936	0.000	161.345
<hr/>							
Standard Deviations:							
L2 : SD(Intercept)	1.356	0.090	1.198	1.549	---	---	5041.908
Residual SD	1.143	0.016	1.113	1.175	---	---	8223.888
<hr/>							
Standardized Coefficients:							
SleepQual	0.151	0.013	0.127	0.176	142.389	0.000	1996.211
SleepQual.mean[Person]	0.412	0.055	0.297	0.513	55.809	0.000	169.087
<hr/>							
Proportion Variance Explained							
by Coefficients	0.193	0.044	0.112	0.286	---	---	175.664
by Level-2 Random Intercepts	0.471	0.042	0.392	0.555	---	---	414.230
by Level-1 Residual Variation	0.334	0.027	0.280	0.387	---	---	511.829

# 95% CREDIBLE INTERVALS

---

- The 95% credible interval gives limits spanning 95% of the parameter's range
- Akin to a confidence interval, but references a range of highly plausible parameter values for one data set



# BLIMP OUTPUT

---

Outcome Variable: PosAffect

Group Mean Centered: SleepQual

Parameters	Estimate	StdDev	2.5%	97.5%	ChiSq	PValue	N_Eff
<hr/>							
Variances:							
L2 : Var(Intercept)	1.839	0.246	1.436	2.399	---	---	5051.168
Residual Var.	1.307	0.037	1.238	1.380	---	---	8227.580
<hr/>							
Coefficients:							
Intercept	4.997	0.133	4.761	5.277	1423.604	0.000	154.295
SleepQual	0.173	0.013	0.147	0.199	174.595	0.000	9302.345
SleepQual.mean[Person]	0.598	0.094	0.416	0.788	40.936	0.000	161.345
<hr/>							
Standard Deviations:							
L2 : SD(Intercept)	1.356	0.090	1.198	1.549	---	---	5041.908
Residual SD	1.143	0.016	1.113	1.175	---	---	8223.888
<hr/>							
Standardized Coefficients:							
SleepQual	0.151	0.013	0.127	0.176	142.389	0.000	1996.211
SleepQual.mean[Person]	0.412	0.055	0.297	0.513	55.809	0.000	169.087
<hr/>							
Proportion Variance Explained							
by Coefficients	0.193	0.044	0.112	0.286	---	---	175.664
by Level-2 Random Intercepts	0.471	0.042	0.392	0.555	---	---	414.230
by Level-1 Residual Variation	0.334	0.027	0.280	0.387	---	---	511.829

# MCMC VS. ML COMPARISON

---

- ML and MCMC estimates are effectively equivalent

Parameter	REML		FIML		Bayesian MCMC	
	Est.	Std. Err.	Est.	Std. Err.	Est.	Std. Err.
Fixed intercept	5.05	0.12	5.05	0.12	5.00	0.13
Sleep (within-person)	0.17	0.01	0.17	0.01	0.17	0.01
Sleep (between-person)	0.58	0.09	0.58	0.08	0.60	0.09
Random intercept variance	1.83	--	1.80	--	1.84	0.25
Residual within-person variance	1.31	--	1.30	--	1.31	0.04

# MCMC AS COMPUTATIONAL FREQUENTISM

---

- Several recent innovations in the multilevel modeling literature (e.g., “dynamic” models with lagged effects) are available only via MCMC estimation
- This has led to a perspective in which MCMC is used for computational purposes rather than philosophical appeal (computational frequentism; Levy & McNeish, 2021)
- Essentially, MCMC results are surrogates for frequentist point estimates, standard errors, and test statistics

# BLIMP OUTPUT

---

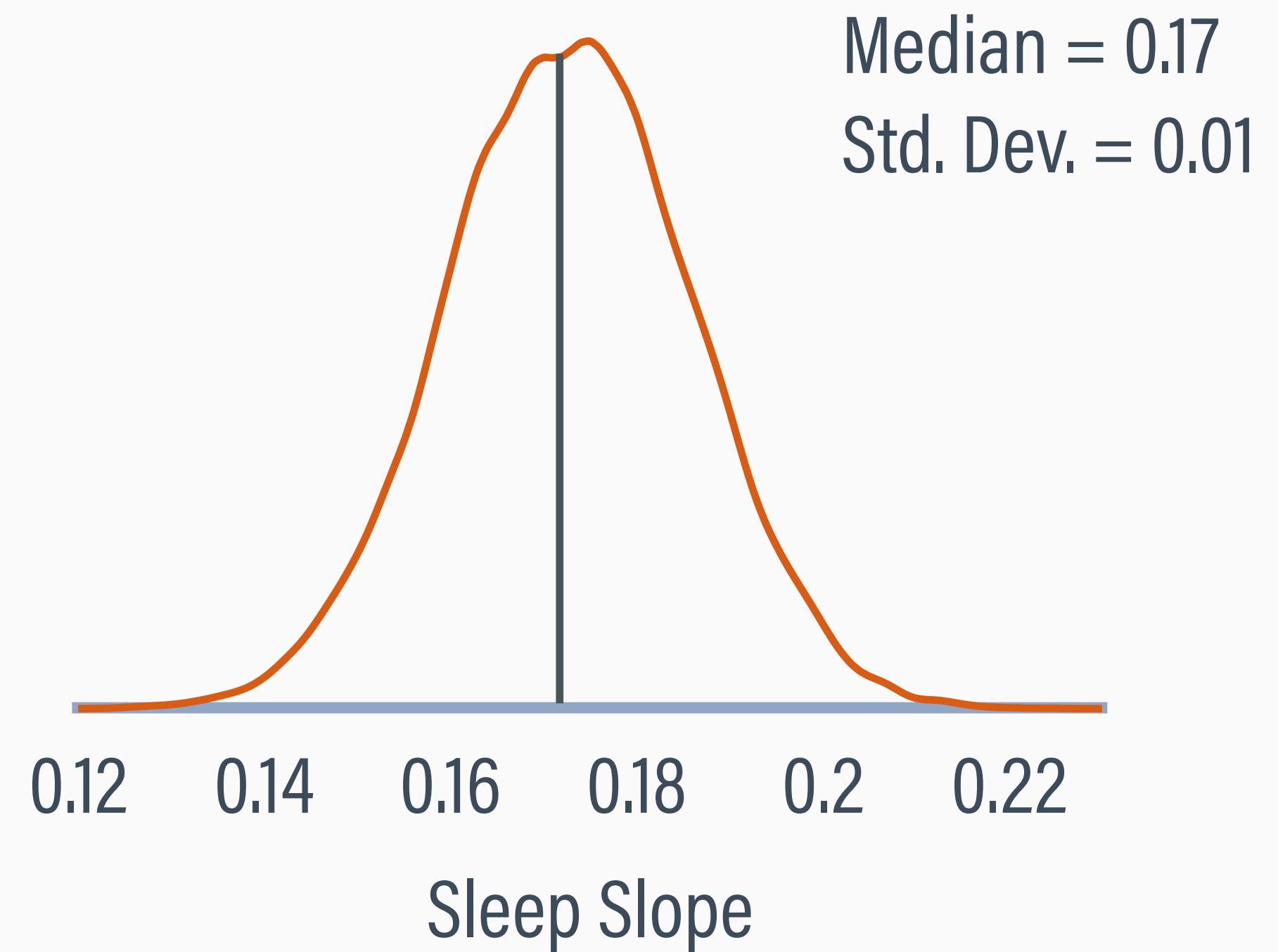
Outcome Variable: PosAffect

Group Mean Centered: SleepQual

Parameters	Estimate	StdDev	2.5%	97.5%	ChiSq	PValue	N_Eff
<hr/>							
Variances:							
L2 : Var(Intercept)	1.839	0.246	1.436	2.399	---	---	5051.168
Residual Var.	1.307	0.037	1.238	1.380	---	---	8227.580
<hr/>							
Coefficients:							
Intercept	4.997	0.133	4.761	5.277	1423.604	0.000	154.295
SleepQual	0.173	0.013	0.147	0.199	174.595	0.000	9302.345
SleepQual.mean[Person]	0.598	0.094	0.416	0.788	40.936	0.000	161.345
<hr/>							
Standard Deviations:							
L2 : SD(Intercept)	1.356	0.090	1.198	1.549	---	---	5041.908
Residual SD	1.143	0.016	1.113	1.175	---	---	8223.888
<hr/>							
Standardized Coefficients:							
SleepQual	0.151	0.013	0.127	0.176	142.389	0.000	1996.211
SleepQual.mean[Person]	0.412	0.055	0.297	0.513	55.809	0.000	169.087
<hr/>							
Proportion Variance Explained							
by Coefficients	0.193	0.044	0.112	0.286	---	---	175.664
by Level-2 Random Intercepts	0.471	0.042	0.392	0.555	---	---	414.230
by Level-1 Residual Variation	0.334	0.027	0.280	0.387	---	---	511.829

# ILLUSTRATION: WITHIN-CLUSTER SLOPE

- The most likely population slope for these data (the posterior median) is  $\beta_1 = 0.17$ , and the parameter's standard deviation is 0.01
- From a computational frequentism lens, these can serve as MCMC-generated point estimates and standard errors



# BLIMP OUTPUT

---

Outcome Variable: PosAffect

Group Mean Centered: SleepQual

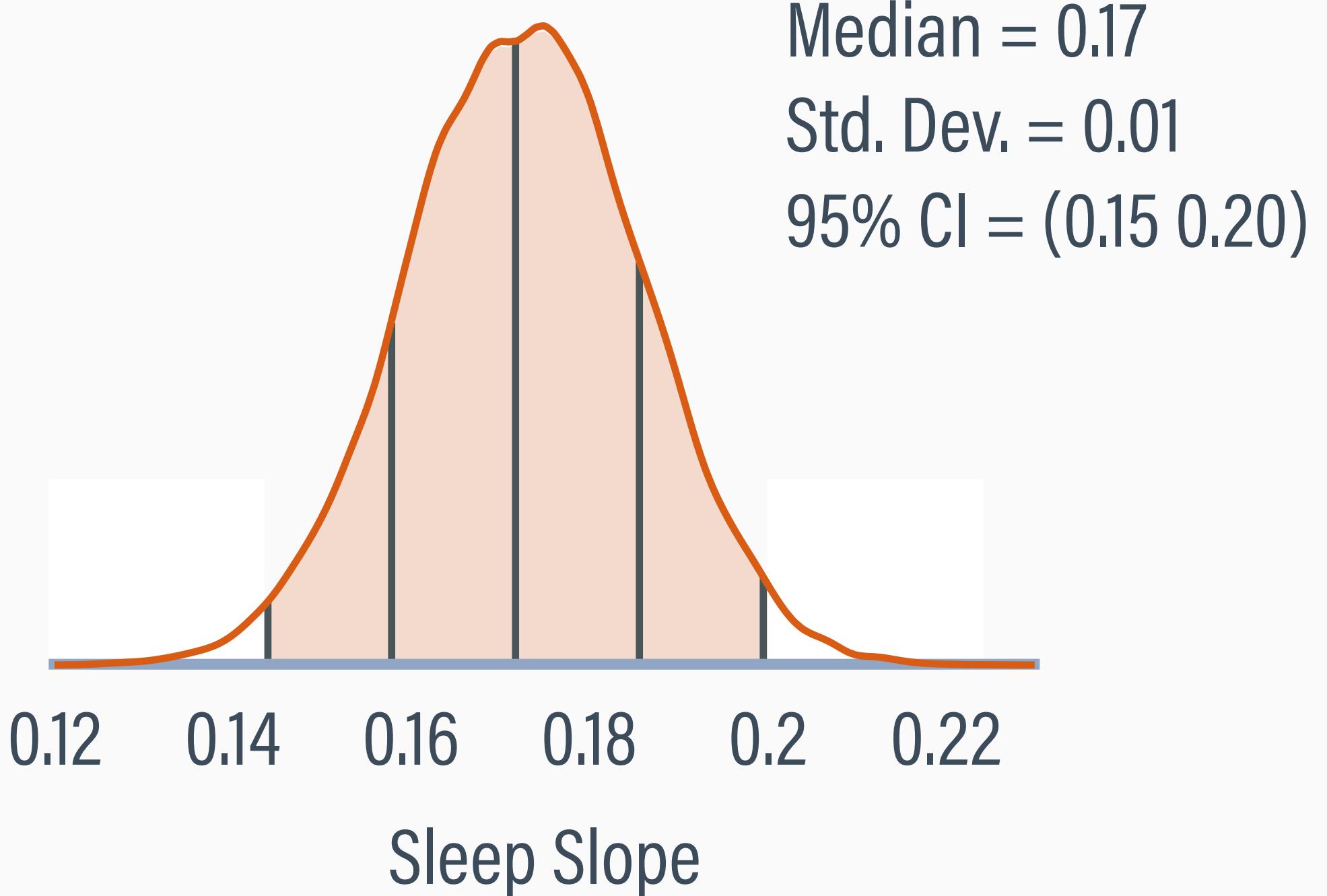
Parameters	Estimate	StdDev	2.5%	97.5%	ChiSq	PValue	N_Eff
<hr/>							
Variances:							
L2 : Var(Intercept)	1.839	0.246	1.436	2.399	---	---	5051.168
Residual Var.	1.307	0.037	1.238	1.380	---	---	8227.580
<hr/>							
Coefficients:							
Intercept	4.997	0.133	4.761	5.277	1423.604	0.000	154.295
SleepQual	<b>0.173</b>	<b>0.013</b>	0.147	0.199	174.595	0.000	9302.345
SleepQual.mean[Person]	0.598	0.094	0.416	0.788	40.936	0.000	161.345
<hr/>							
Standard Deviations:							
L2 : SD(Intercept)	1.356	0.090	1.198	1.549	---	---	5041.908
Residual SD	1.143	0.016	1.113	1.175	---	---	8223.888
<hr/>							
Standardized Coefficients:							
SleepQual	0.151	0.013	0.127	0.176	142.389	0.000	1996.211
SleepQual.mean[Person]	0.412	0.055	0.297	0.513	55.809	0.000	169.087
<hr/>							
Proportion Variance Explained							
by Coefficients	0.193	0.044	0.112	0.286	---	---	175.664
by Level-2 Random Intercepts	0.471	0.042	0.392	0.555	---	---	414.230
by Level-1 Residual Variation	0.334	0.027	0.280	0.387	---	---	511.829

---

# ILLUSTRATION, CONTINUED

---

- 95% of the plausible parameter values that could have produced these data range between 0.15 and 0.20
- From a computational frequentism lens, these can serve as MCMC-generated confidence interval limits



# BLIMP OUTPUT

---

Outcome Variable: PosAffect

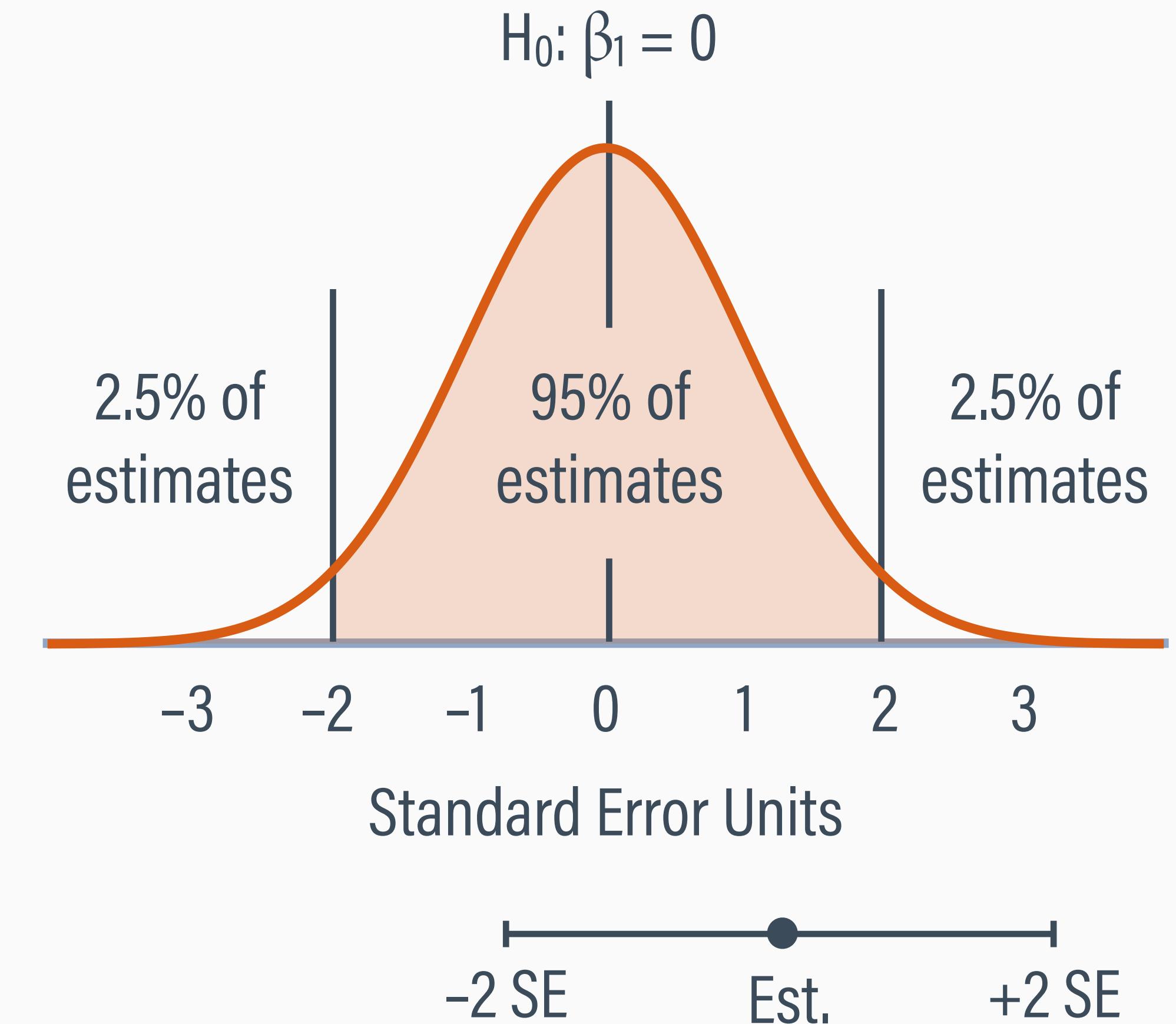
Group Mean Centered: SleepQual

Parameters	Estimate	StdDev	2.5%	97.5%	ChiSq	PValue	N_Eff
<hr/>							
Variances:							
L2 : Var(Intercept)	1.839	0.246	1.436	2.399	---	---	5051.168
Residual Var.	1.307	0.037	1.238	1.380	---	---	8227.580
<hr/>							
Coefficients:							
Intercept	4.997	0.133	4.761	5.277	1423.604	0.000	154.295
SleepQual	0.173	0.013	0.147	0.199	174.595	0.000	9302.345
SleepQual.mean[Person]	0.598	0.094	0.416	0.788	40.936	0.000	161.345
<hr/>							
Standard Deviations:							
L2 : SD(Intercept)	1.356	0.090	1.198	1.549	---	---	5041.908
Residual SD	1.143	0.016	1.113	1.175	---	---	8223.888
<hr/>							
Standardized Coefficients:							
SleepQual	0.151	0.013	0.127	0.176	142.389	0.000	1996.211
SleepQual.mean[Person]	0.412	0.055	0.297	0.513	55.809	0.000	169.087
<hr/>							
Proportion Variance Explained							
by Coefficients	0.193	0.044	0.112	0.286	---	---	175.664
by Level-2 Random Intercepts	0.471	0.042	0.392	0.555	---	---	414.230
by Level-1 Residual Variation	0.334	0.027	0.280	0.387	---	---	511.829

---

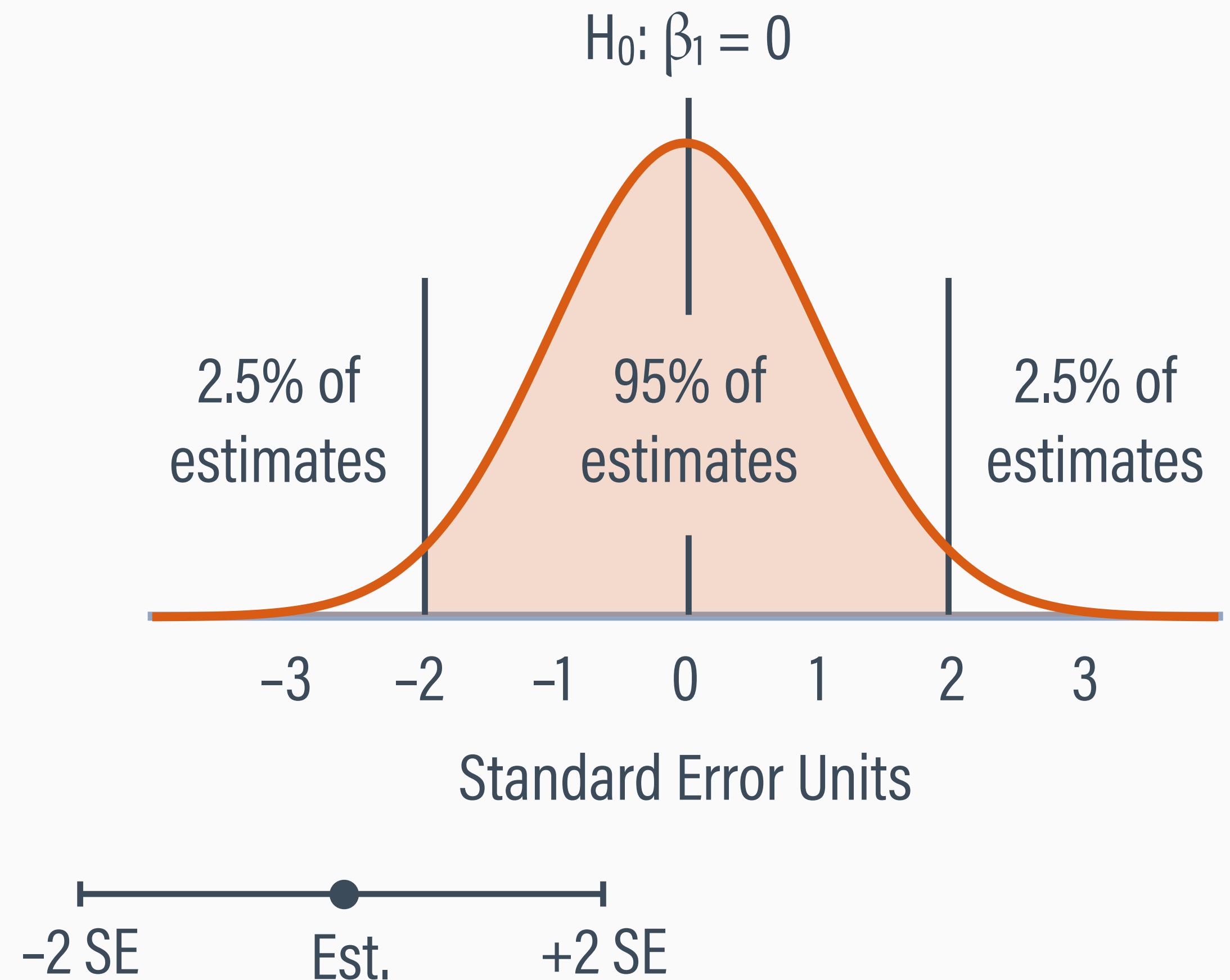
# SIGNIFICANCE TESTING VIA INTERVALS

- A confidence interval applies the  $\pm 2$  standard error band to an estimate
- If the 95% confidence interval includes zero, the estimate cannot be in the critical region, and  $p > .05$



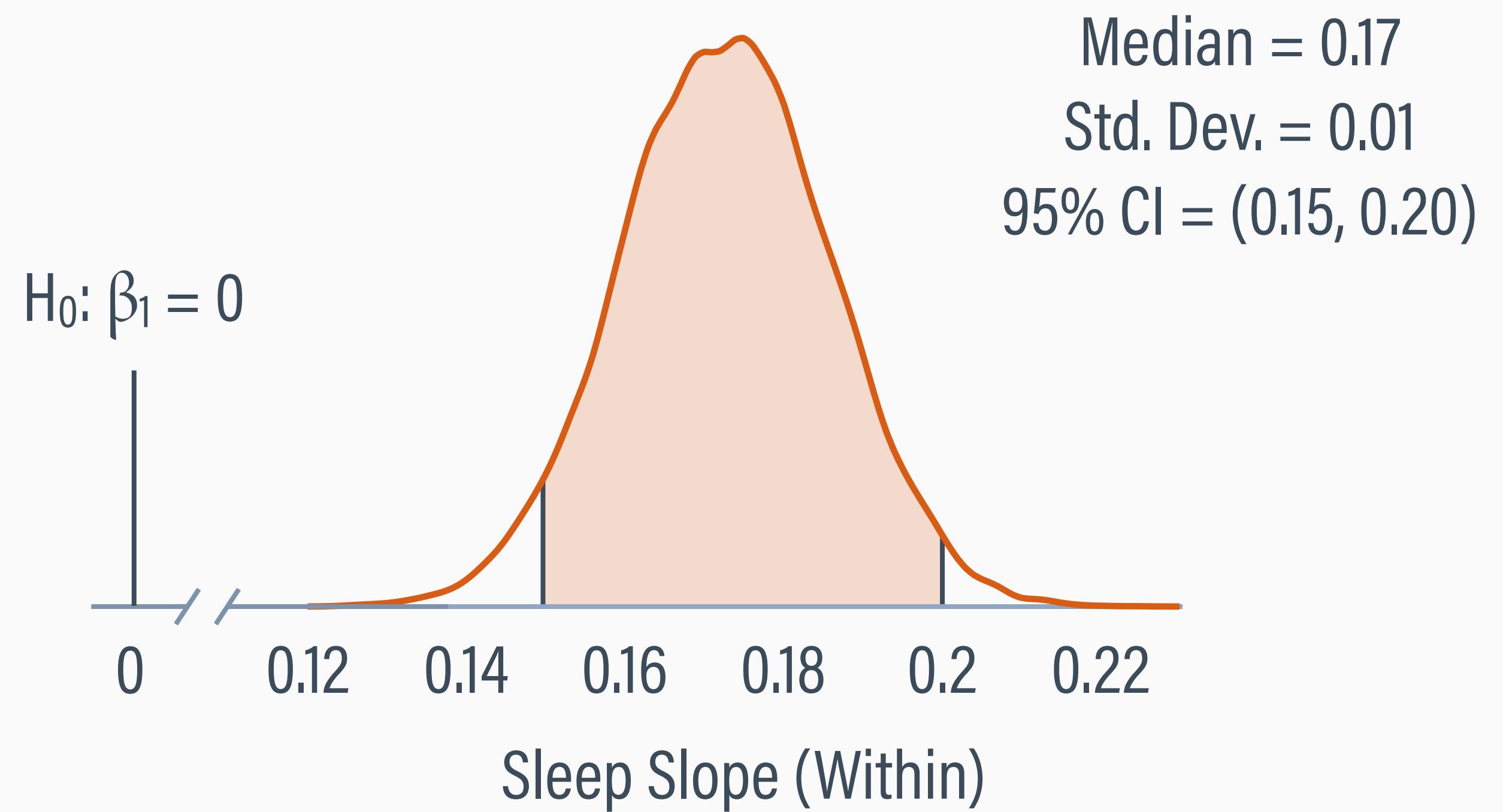
# SIGNIFICANCE TESTING VIA INTERVALS

- A confidence interval applies the  $\pm 2$  standard error band to an estimate
- If the 95% confidence interval does not include zero, the estimate must be in the critical region, and  $p < .05$



# SIGNIFICANCE TEST ILLUSTRATION

- The 95% credible interval (0.15, 0.20) does not include zero, so we refute the null hypothesis ( $p < .05$ )
- A population slope equal to zero is unlikely to have produced these data



# FREQUENTIST WALD TEST

---

- The Wald chi-square is an alternate test statistic that equals the square of the z-statistic (or t-test)

$$\chi^2_{\text{ML}} = \frac{(\hat{\theta} - \theta_0)^2}{\text{SE}^2} = \frac{(\text{estimate} - \text{null})^2}{(\text{standard error})^2} = z^2$$

- Multivariate versions of the Wald test can evaluate multiple parameters simultaneously

# MCMC WALD TEST

---

- The MCMC-based Wald test (Asparouhov & Muthén, 2021) replaces the point estimate and standard error with the parameter's mean and standard deviation

$$\chi^2_{\text{MCMC}} = \frac{(\theta - \theta_0)^2}{\text{SD}^2} = \frac{(\text{posterior mean} - \text{null})^2}{(\text{posterior standard deviation})^2} = z^2$$

- MCMC quantities replace FIML estimates in the test statistic

# BLIMP OUTPUT

---

Outcome Variable: PosAffect

Group Mean Centered: SleepQual

Parameters	Estimate	StdDev	2.5%	97.5%	ChiSq	PValue	N_Eff
<hr/>							
Variances:							
L2 : Var(Intercept)	1.839	0.246	1.436	2.399	---	---	5051.168
Residual Var.	1.307	0.037	1.238	1.380	---	---	8227.580
<hr/>							
Coefficients:							
Intercept	4.997	0.133	4.761	5.277	1423.604	0.000	154.295
SleepQual	0.173	0.013	0.147	0.199	174.595	0.000	9302.345
SleepQual.mean[Person]	0.598	0.094	0.416	0.788	40.936	0.000	161.345
<hr/>							
Standard Deviations:							
L2 : SD(Intercept)	1.356	0.090	1.198	1.549	---	---	5041.908
Residual SD	1.143	0.016	1.113	1.175	---	---	8223.888
<hr/>							
Standardized Coefficients:							
SleepQual	0.151	0.013	0.127	0.176	142.389	0.000	1996.211
SleepQual.mean[Person]	0.412	0.055	0.297	0.513	55.809	0.000	169.087
<hr/>							
Proportion Variance Explained							
by Coefficients	0.193	0.044	0.112	0.286	---	---	175.664
by Level-2 Random Intercepts	0.471	0.042	0.392	0.555	---	---	414.230
by Level-1 Residual Variation	0.334	0.027	0.280	0.387	---	---	511.829

---

# WALD TEST INTERPRETATION

---

- The test refutes the null hypothesis that the population within-person regression slopes equals zero
- $\chi^2(1) = 174.60, p < .001$
- If the null is true in the population, a slope as large as the one from the data would result in fewer than 1 out of 1000 random samples (this is a 100% frequentist interpretation!)

# OUTLINE

- 1 Analysis Example
- 2 Frequentist vs. Bayesian Statistical Paradigms
- 3 Maximum Likelihood Estimation
- 4 MCMC Estimation
- 5 MCMC Diagnostics
- 6 Checking MLM Assumptions

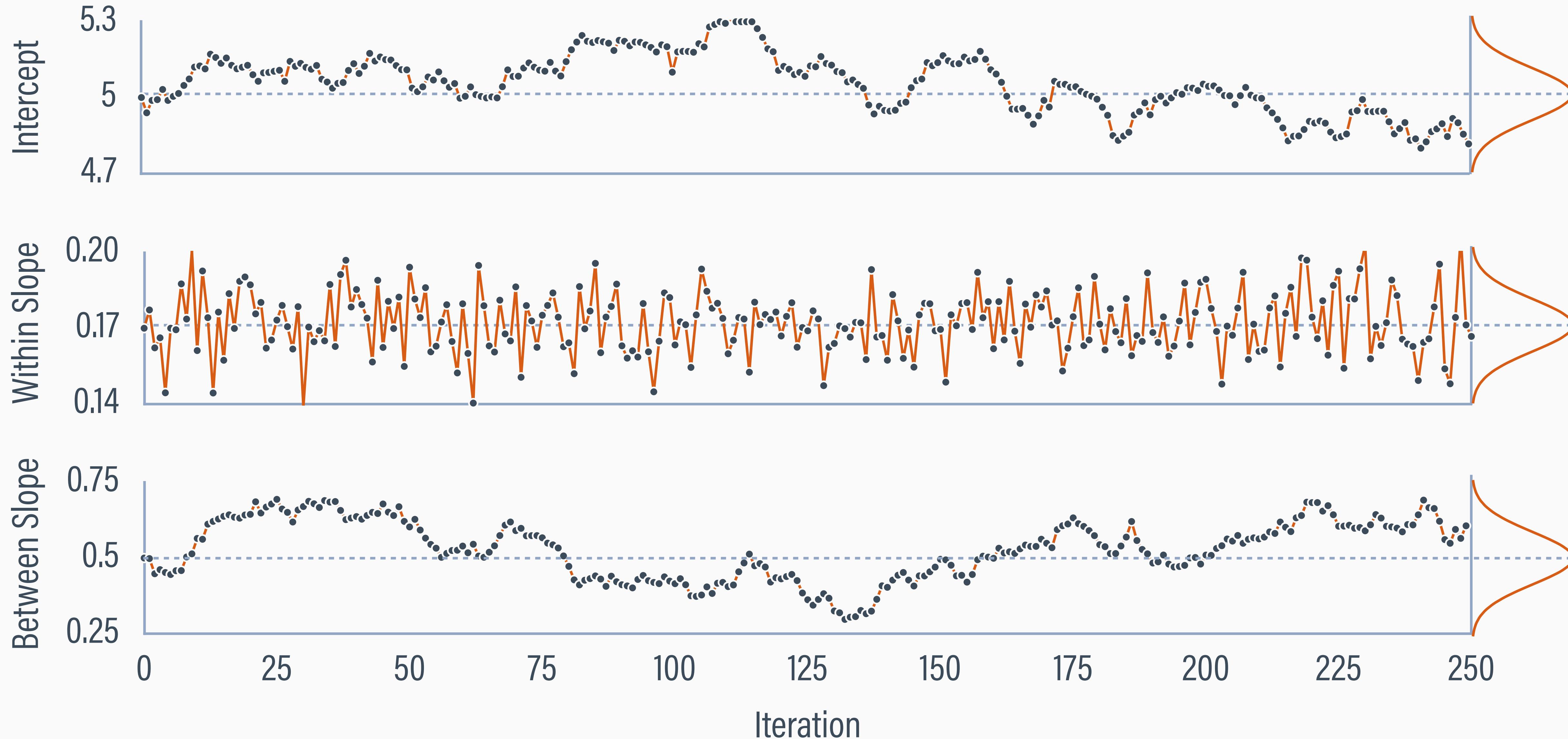
# MCMC ALGORITHM

---

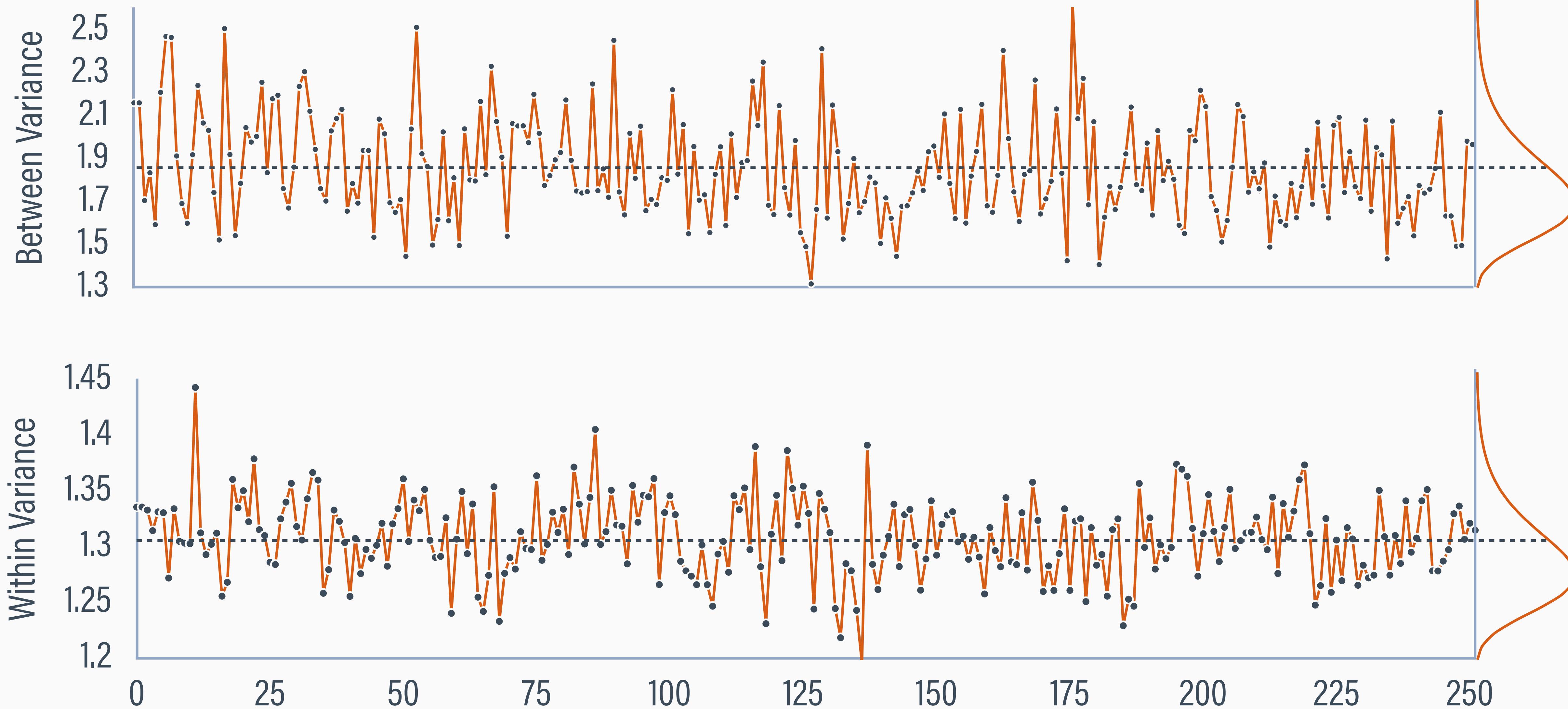


- » Do for  $t = 1$  to  $T$  iterations
  - » Estimate regression coefficients
  - » Estimate random intercept residuals
  - » Estimate level-2 intercept variance
  - » Estimate level-1 residual variance
  - » Impute missing data
- » Repeat

# COEFFICIENTS FROM 250 MCMC CYCLES



# VARIANCES FROM 250 MCMC CYCLES



# MCMC CONVERGENCE

---

- MCMC parameter values continually fluctuate across cycles
- MCMC converges when posterior distributions are stationary, meaning that parameter values oscillate around a stable mean, and their variation doesn't change with additional iterations
- MCMC should “warm up” until convergence, after which parameter values are saved for the distributional summaries

# MCMC ALGORITHM SETTINGS

---

**BURN:** 10000;

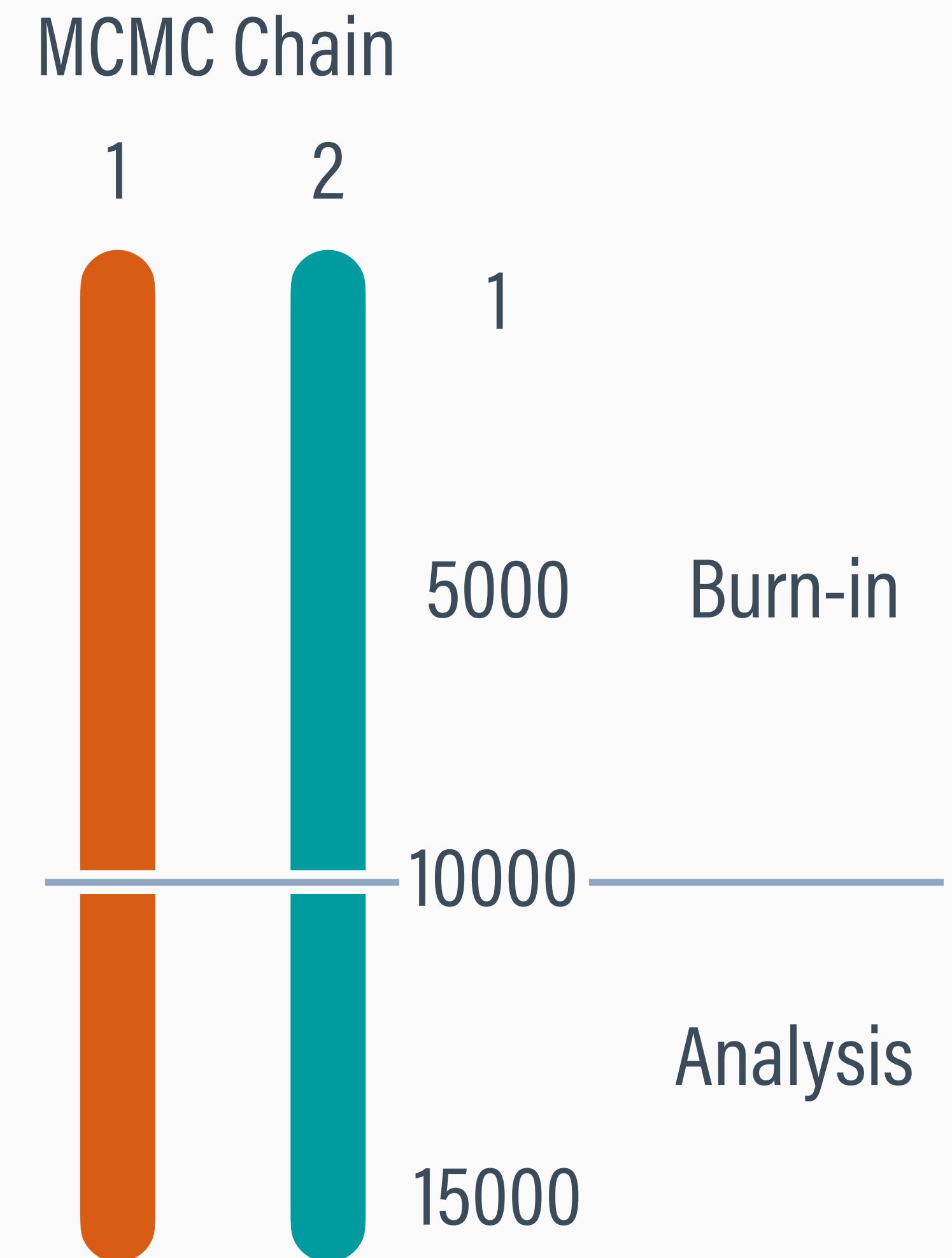
# warm up iterations (per chain)

**ITERATIONS:** 10000;

# analysis iterations (both chains)

**SEED:** 90291;

# random number seed



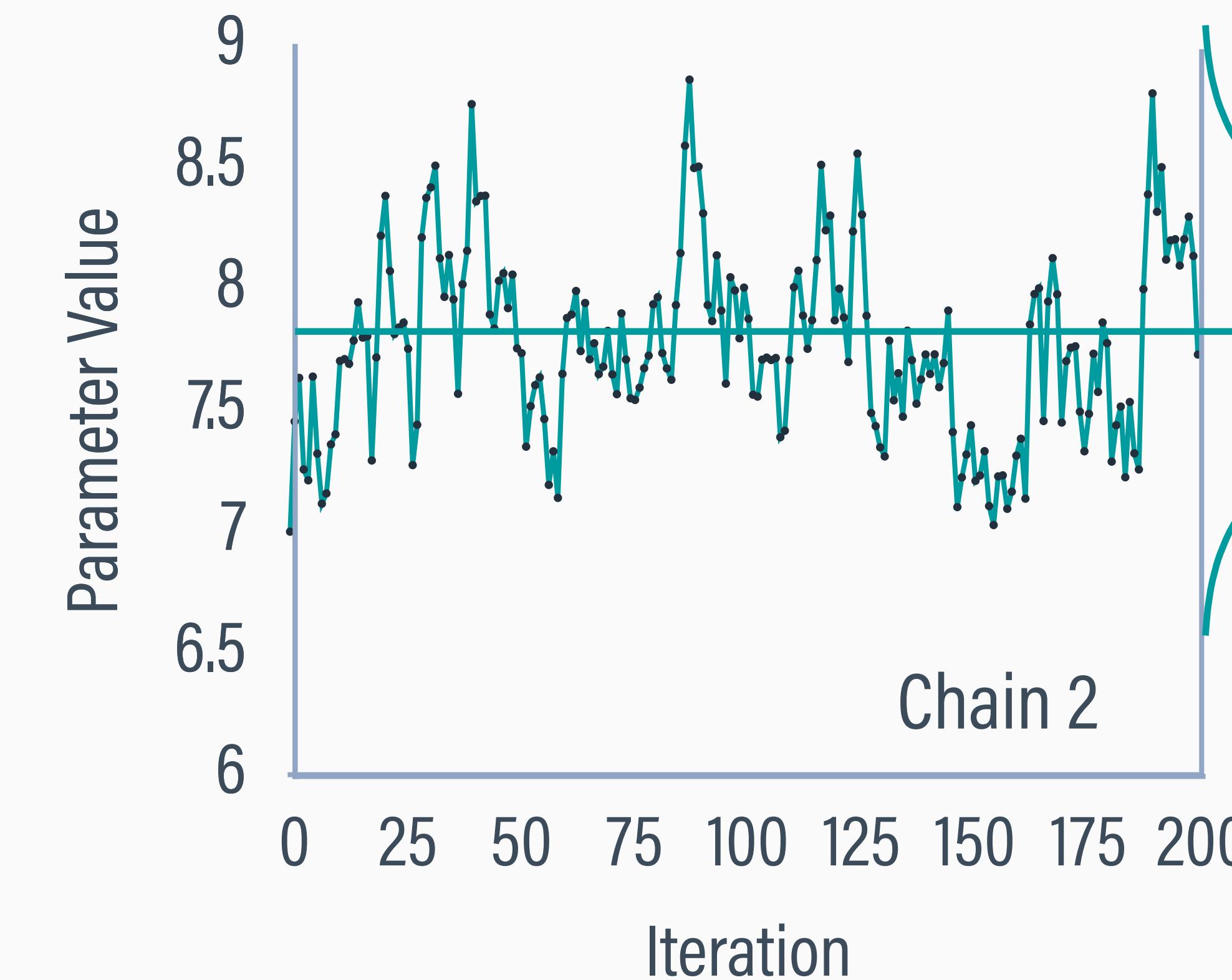
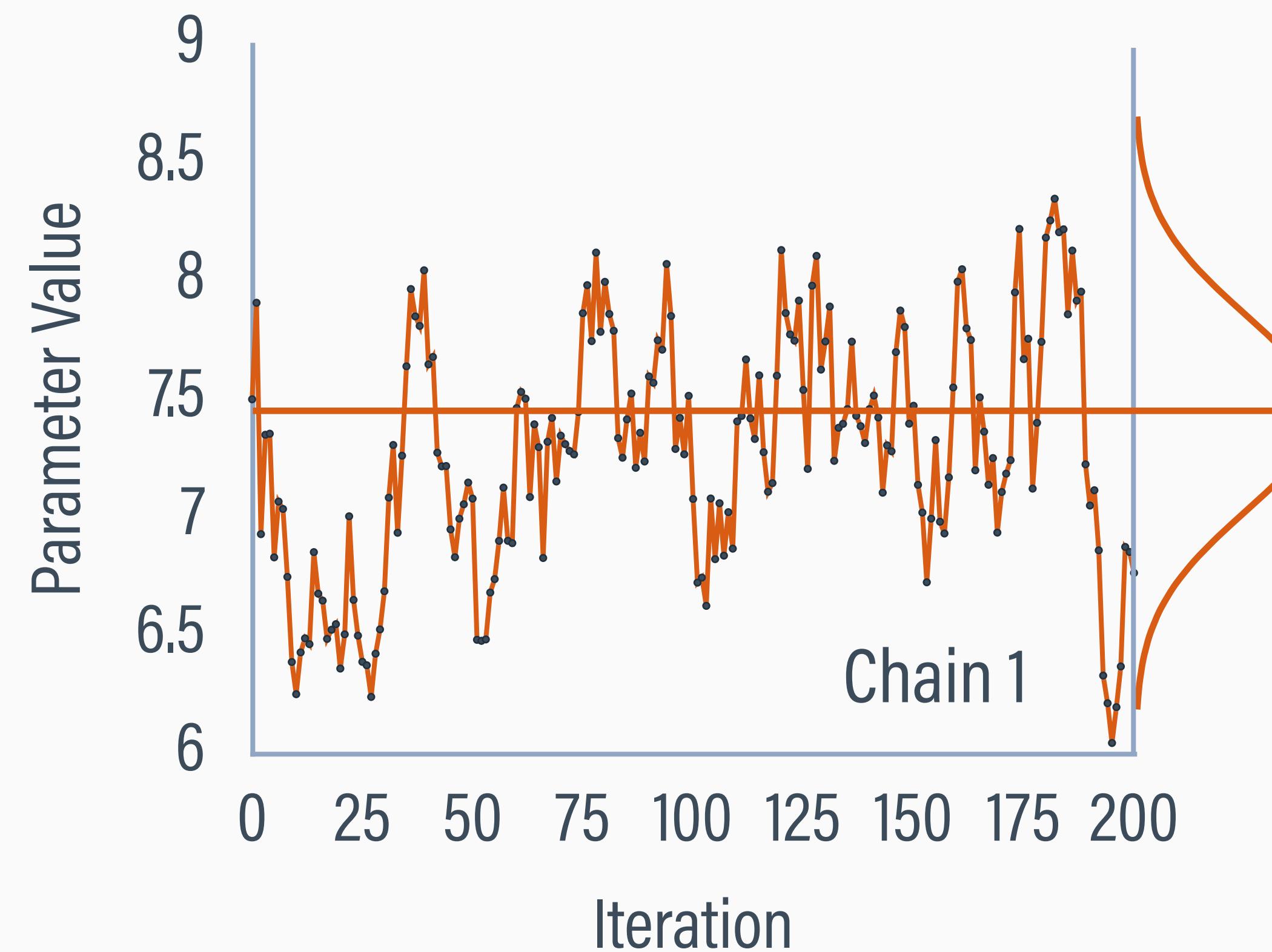
# POTENTIAL SCALE REDUCTION FACTOR

---

- The potential scale reduction factor (PSRF) compares the similarity of parameters generated from two MCMC processes
- MCMC converges when the two unique processes (chains) give parameter values with same mean and spread
- PSRFs for all parameters should be  $< 1.05$

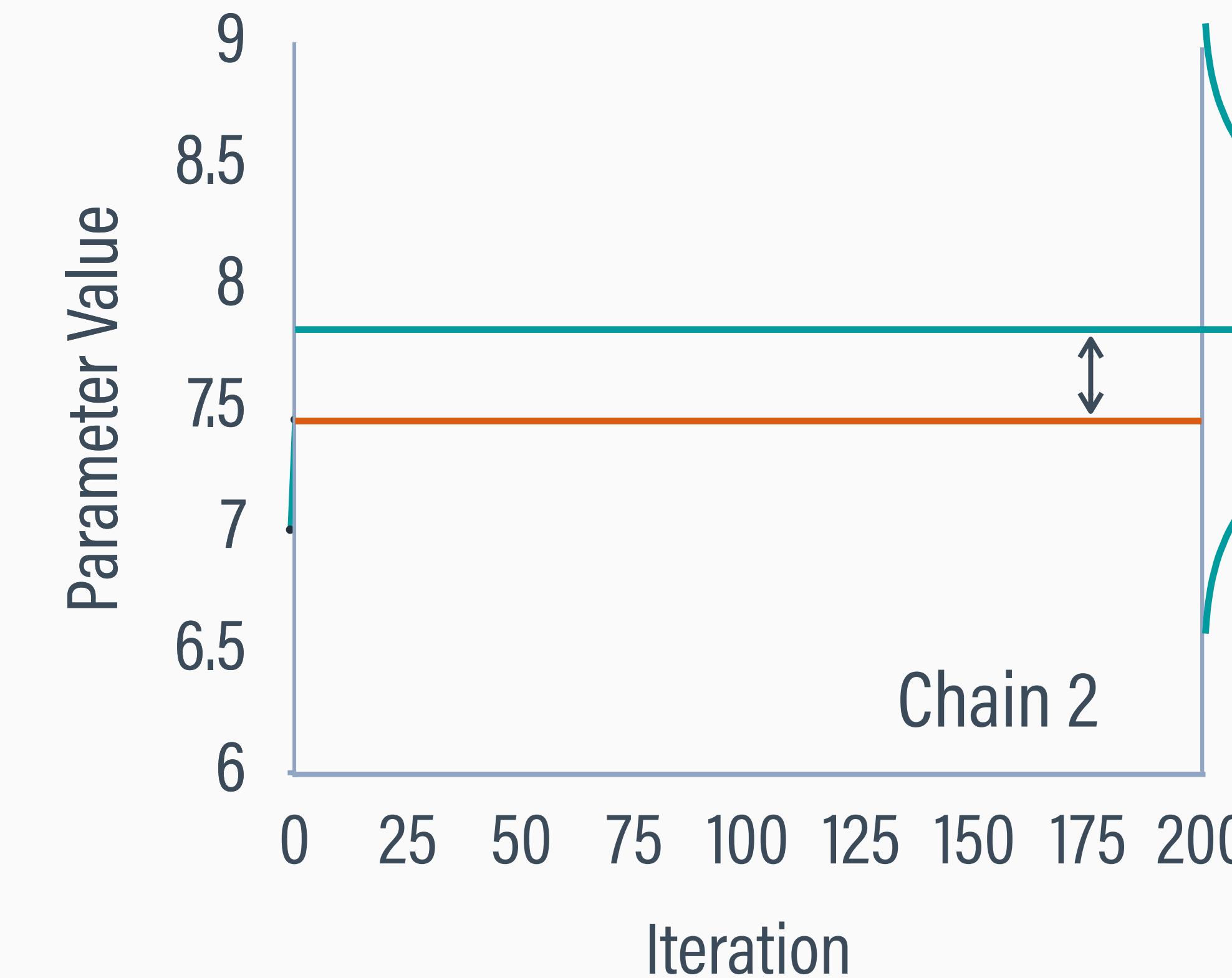
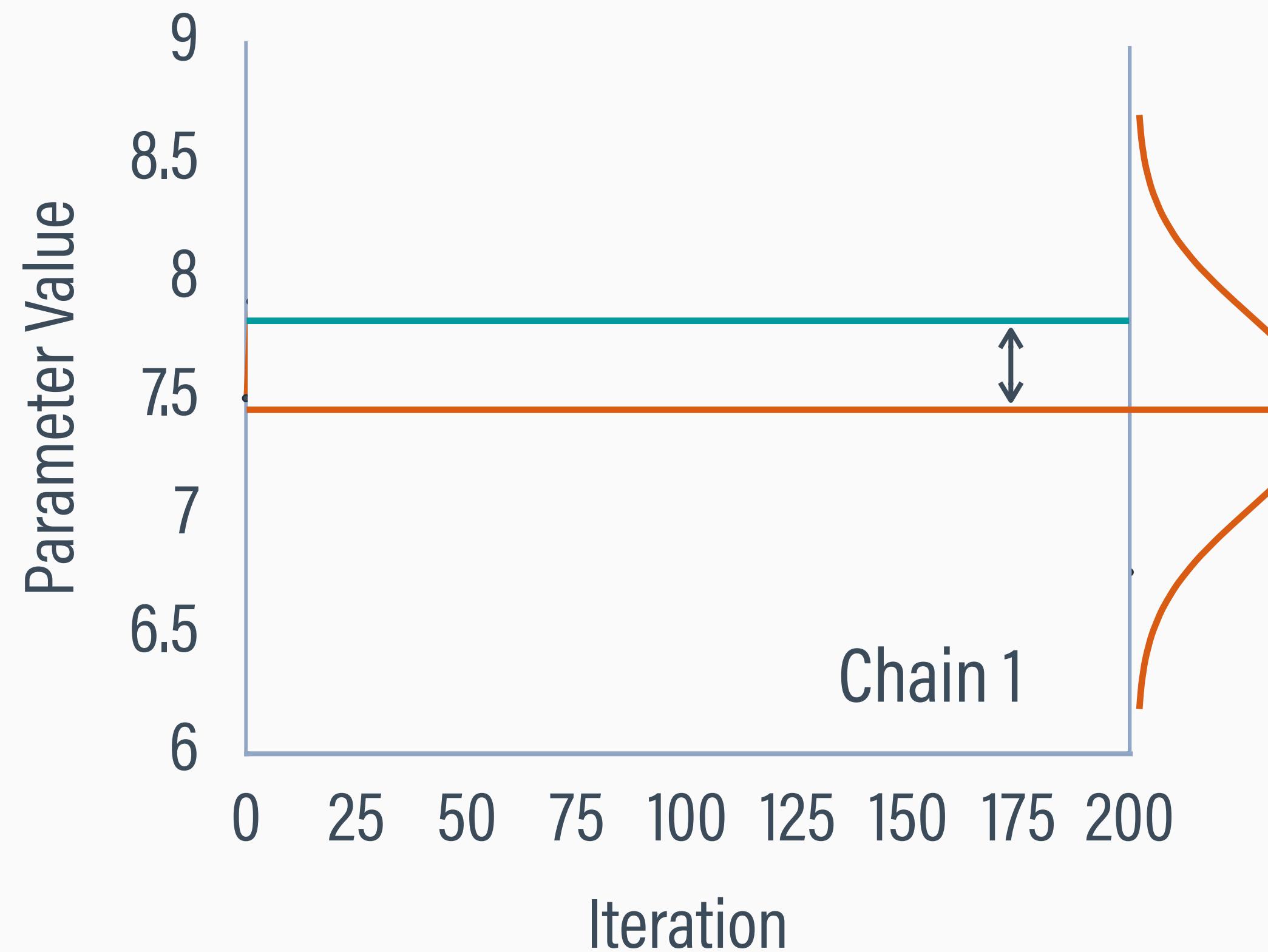
# POTENTIAL SCALE REDUCTION FACTOR

$$\text{PSRF} = \sqrt{\frac{\text{mean difference between chains} + \text{within-chain variation}}{\text{within-chain variation}}}$$



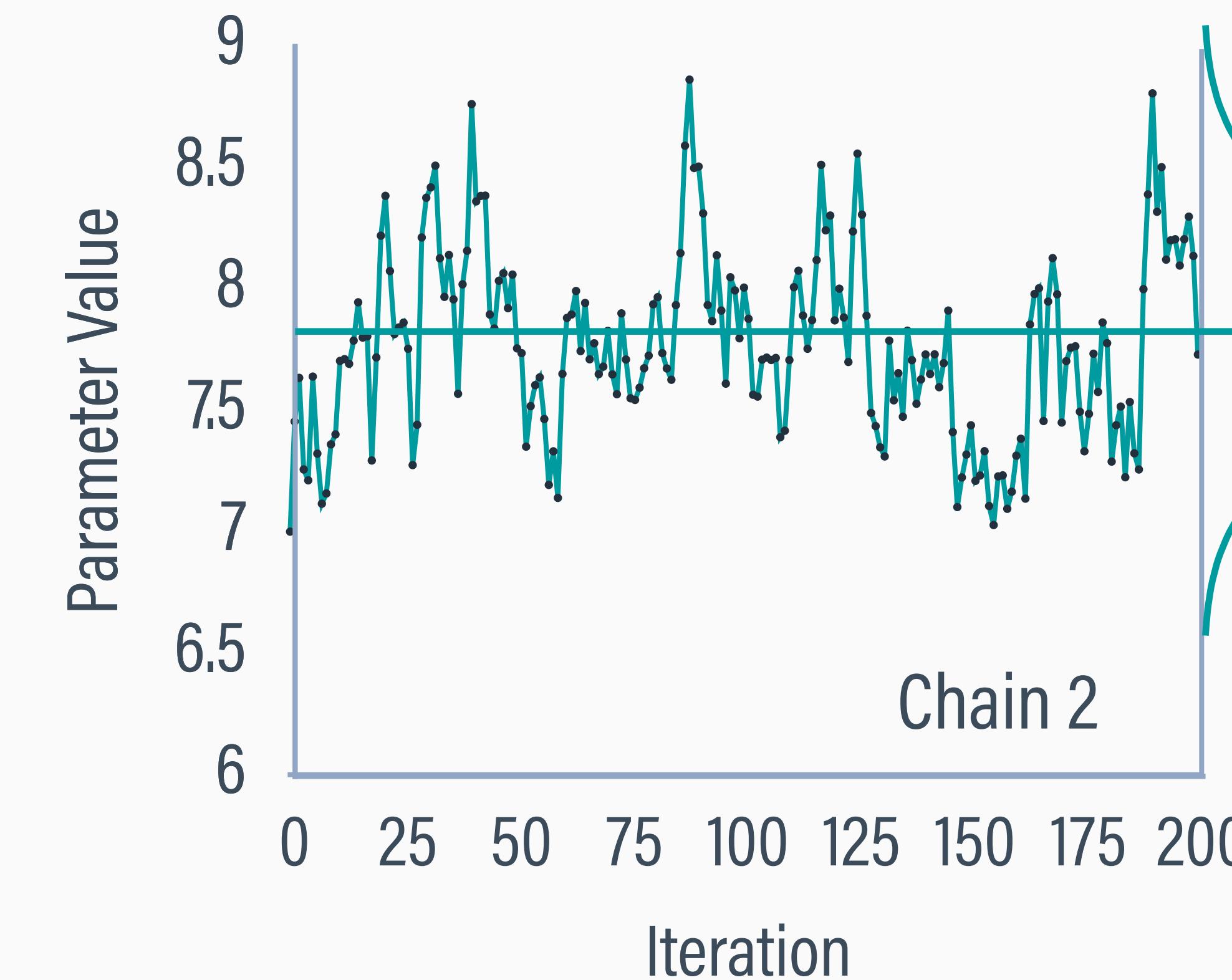
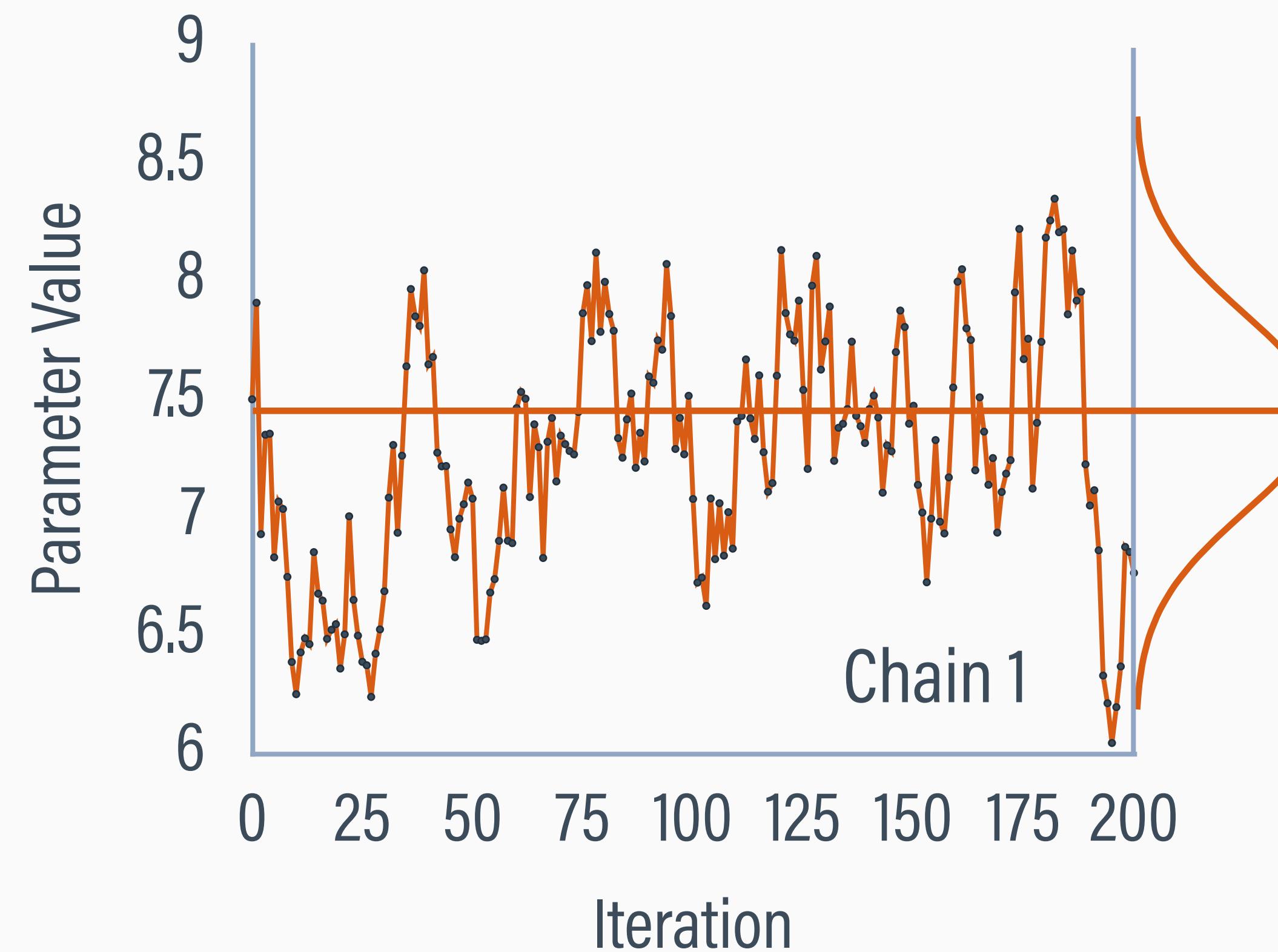
# BETWEEN-CHAIN MEAN DIFFERENCE

$$\text{PSRF} = \sqrt{\frac{\text{mean difference between chains} + \text{within-chain variation}}{\text{within-chain variation}}}$$



# WITHIN-CHAIN VARIATION

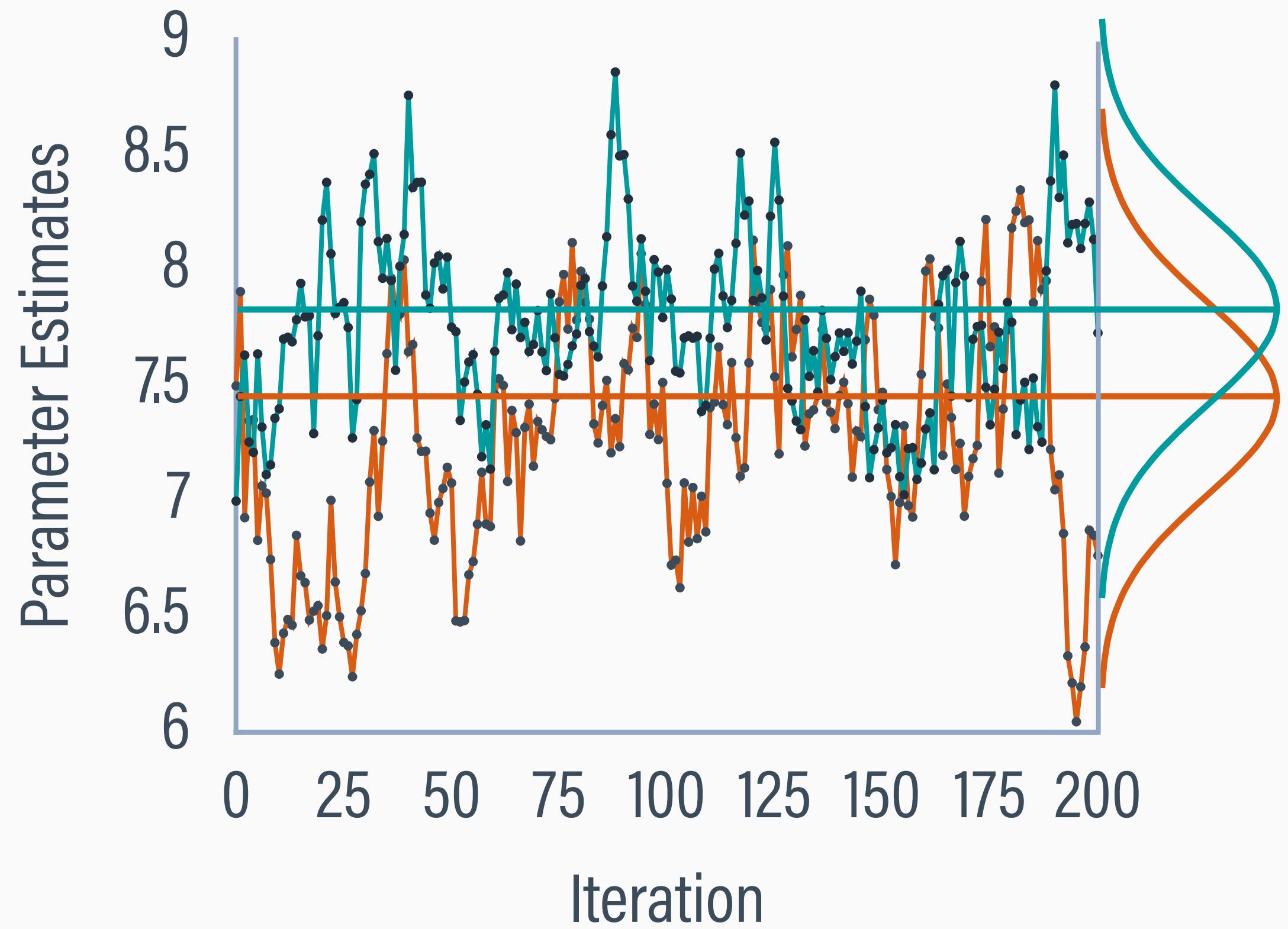
$$\text{PSRF} = \sqrt{\frac{\text{mean difference between chains} + \text{within-chain variation}}{\text{within-chain variation}}}$$



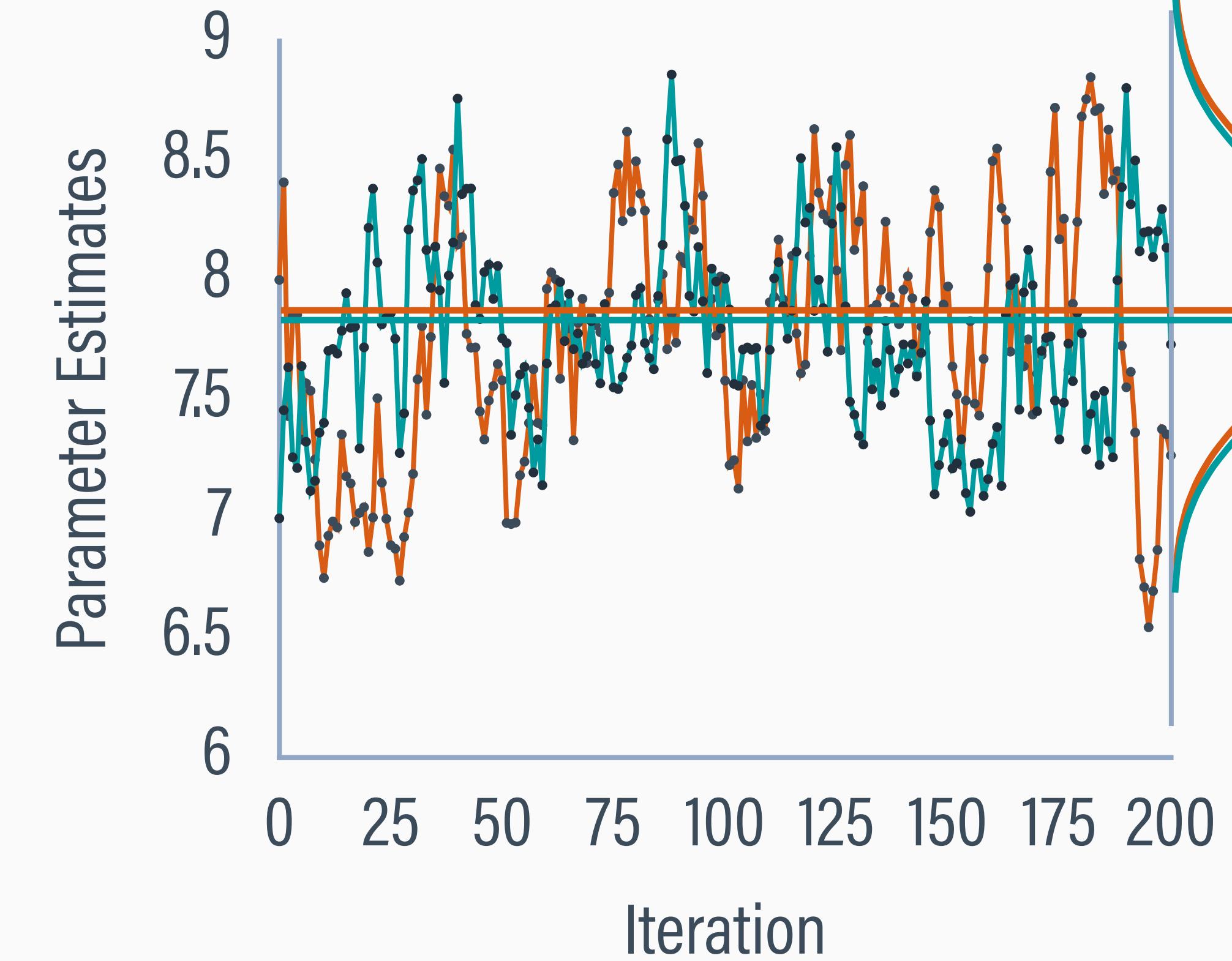
# CONVERGENCE

---

MCMC has not converged because between-chain mean difference is large ( $\text{PSR} > 1.05$ )



MCMC has converged because between-chain mean difference is very small ( $\text{PSR} < 1.05$ )



# PSRF DIAGNOSTIC OUTPUT

---

The number of burn-in iterations is sufficient because the highest PSRF across all parameters is < 1.05 at the end of the burn-in period

## BURN-IN POTENTIAL SCALE REDUCTION (PSR) OUTPUT:

NOTE: Split chain PSR is being used. This splits each chain's iterations to create twice as many chains.

Comparing iterations across 2 chains	Highest PSR	Parameter #
251 to 500	1.428	3
501 to 1000	1.387	3
751 to 1500	1.076	5
1001 to 2000	1.161	11
1251 to 2500	1.064	11
1501 to 3000	1.046	3
1751 to 3500	1.100	11
2001 to 4000	1.029	3
...	...	...
3501 to 7000	1.023	3
3751 to 7500	1.024	3
4001 to 8000	1.006	3
4251 to 8500	1.019	5
4501 to 9000	1.013	5
4751 to 9500	1.012	3
5001 to 10000	» 1.031	3

# CONCLUSIONS

---

- PRSFs indicated that a 10,000-iteration warm-up (burn-in) period was sufficient for achieving convergence
- The highest (worst) PSRF was  $< 1.05$  after  $\sim 4,000$  iterations
- Two independent MCMC chains were producing similar parameters (same mean and spread) after  $\sim 4,000$  iterations
- Always inspect the PSRFs before interpreting results!

# EFFECTIVE SAMPLE SIZE DIAGNOSTIC

---

- Effective sample size reflects the number of independent MCMC samples after accounting for autocorrelation
- Values above ~ 100 are viewed as sufficient, and lower values indicate the need for additional iterations
- Low values (high autocorrelation) indicate that a parameter has weak data support from the data

# BLIMP OUTPUT

## Outcome Variable: PosAffect

# Group Mean Centered: SleepQual

Parameters	Estimate	StdDev	2.5%	97.5%	ChiSq	PValue	N_Eff
<b>Variances:</b>							
L2 : Var(Intercept)	1.839	0.246	1.436	2.399	---	---	5051.168
Residual Var.	1.307	0.037	1.238	1.380	---	---	8227.580
<b>Coefficients:</b>							
Intercept	4.997	0.133	4.761	5.277	1423.604	0.000	154.295
SleepQual	0.173	0.013	0.147	0.199	174.595	0.000	9302.345
SleepQual.mean[Person]	0.598	0.094	0.416	0.788	40.936	0.000	161.345
<b>Standard Deviations:</b>							
L2 : SD(Intercept)	1.356	0.090	1.198	1.549	---	---	5041.908
Residual SD	1.143	0.016	1.113	1.175	---	---	8223.888
...							
<b>Proportion Variance Explained</b>							
by Coefficients	0.193	0.044	0.112	0.286	---	---	175.664
by Level-2 Random Intercepts	0.471	0.042	0.392	0.555	---	---	414.230
by Level-1 Residual Variation	0.334	0.027	0.280	0.387	---	---	511.829

# CONCLUSIONS

---

- Diagnostics show that 10,000 post-warm-up iterations provided adequate precision (all  $N_{eff}$  values  $> 100$ )
- The lowest  $N_{eff}$  values (~154 to 161) occurred for level-2 mean parameters—they rely on the number of clusters rather than the full  $N$ , so they naturally have less support from the data
- Additional iterations would improve these values
- Always inspect the  $N_{eff}$  values before interpreting results!

# OUTLINE

- 1 Analysis Example
- 2 Frequentist vs. Bayesian Statistical Paradigms
- 3 Maximum Likelihood Estimation
- 4 MCMC Estimation
- 5 MCMC Diagnostics
- 6 Checking MLM Assumptions

# MLM ASSUMPTIONS

---

- Associations are linear
- Level-1 residuals are normal with constant variation across level-1 units (days) and level-2 units (persons)
- Level-2 random intercept residuals are normal with constant variation across level-2 units (persons)

$$u_{0j} \sim N(0, \sigma_u^2) \quad \varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$$

# RESIDUAL DIAGNOSTICS

---

- Each MCMC iteration produces different parameter values—and thus different predicted values and residuals
- Residuals from different MCMC steps can be saved to create multiple datasets for plotting
- The `univariate_plot` function (imported at the top of the `rblimp` scripts) graphs residuals from multiple `rblimp` data sets

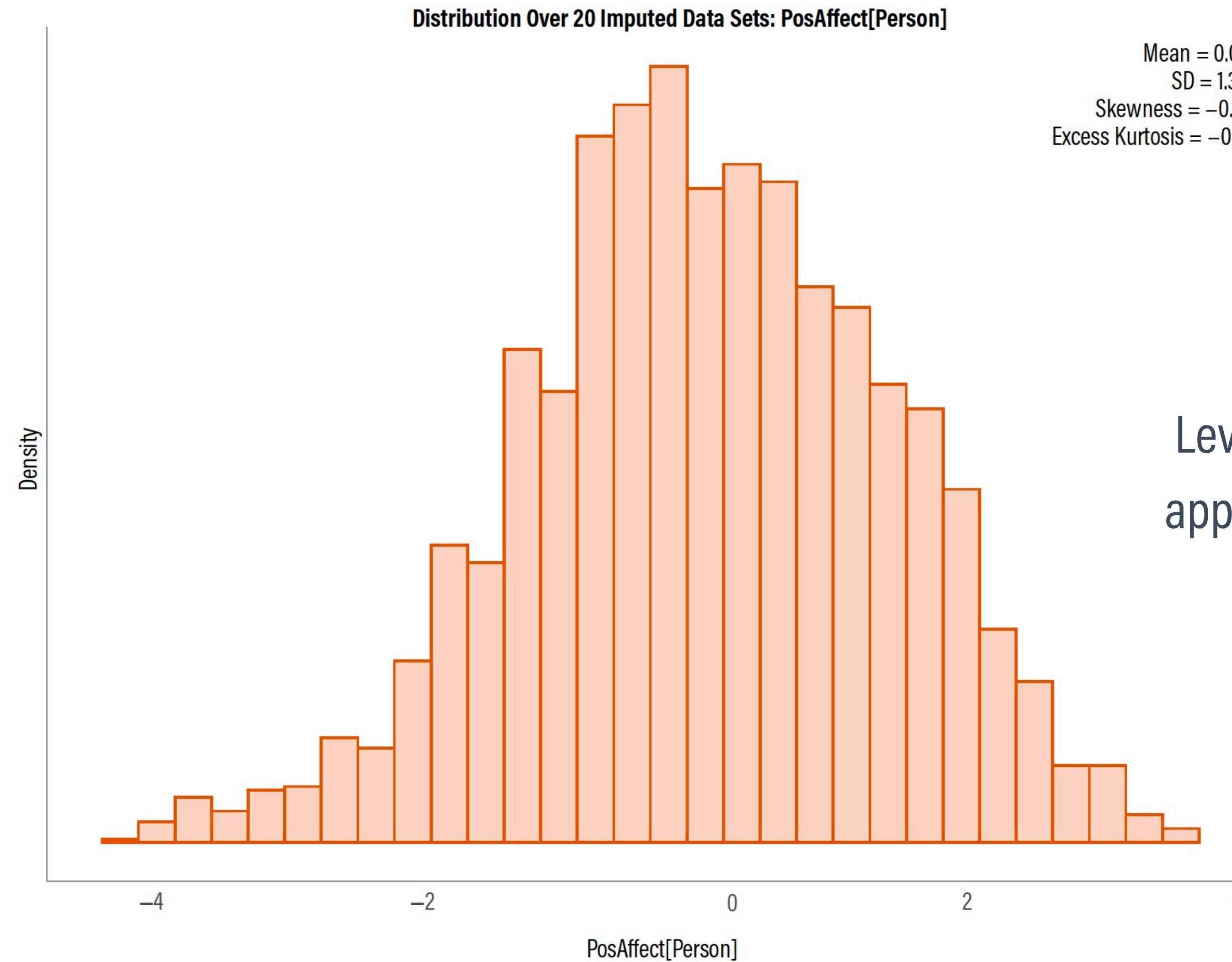
# RBLIMP SCRIPT 3 (MODEL 2)

---

```
model2 <- rblimp(  
  data = PainDiary,  
  clusterid = 'Person',  
  center = 'grandmean = SleepQual.mean; groupmean = SleepQual',  
  model = 'PosAffect ~ intercept SleepQual SleepQual.mean | intercept',  
  seed = 90291,  
  burn = 10000,  
  iter = 10000,  
  nimps = 20)  
  
output(model2)  
univariate_plot(vars = c('PosAffect[Person]', 'PosAffect.residual'), model = model, stats = T)  
bivariate_plot(PosAffect.residual ~ SleepQual, standardize = 'y', model = model2)  
bivariate_plot(PosAffect[Person] ~ SleepQual.mean[Person], standardize = 'y', model = model2)
```

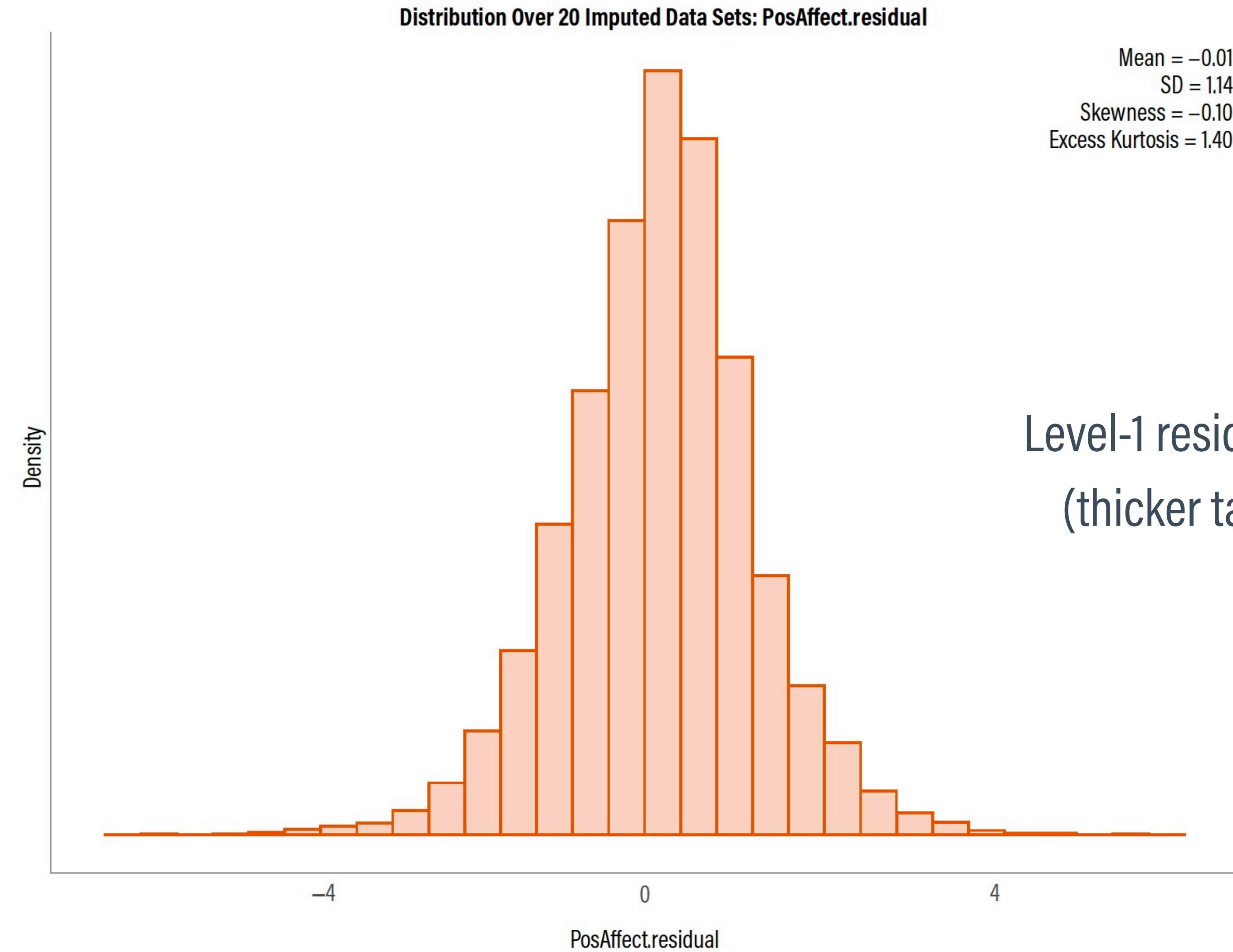
# LEVEL-2 RESIDUALS

---



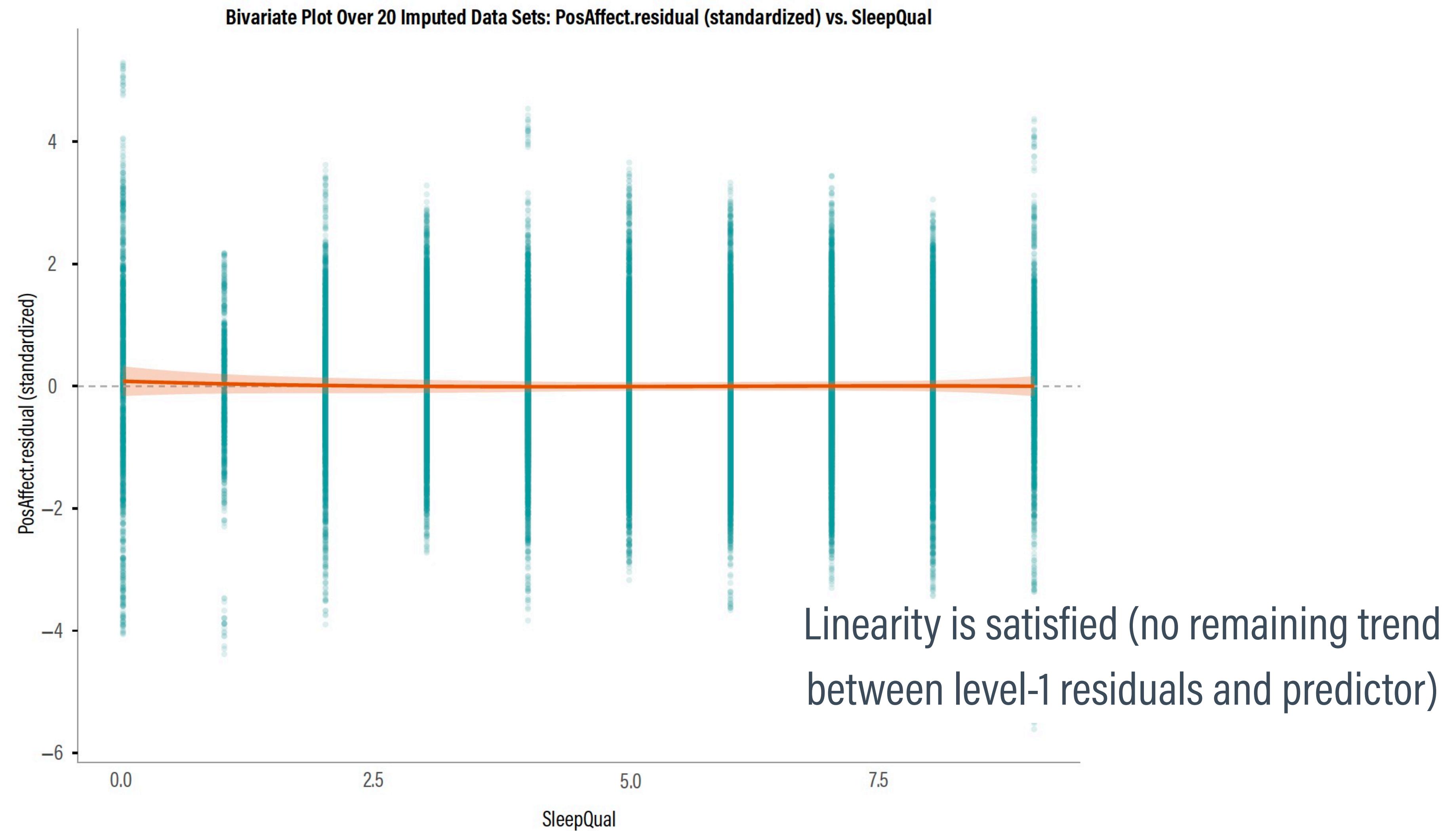
# LEVEL-1 RESIDUALS

---



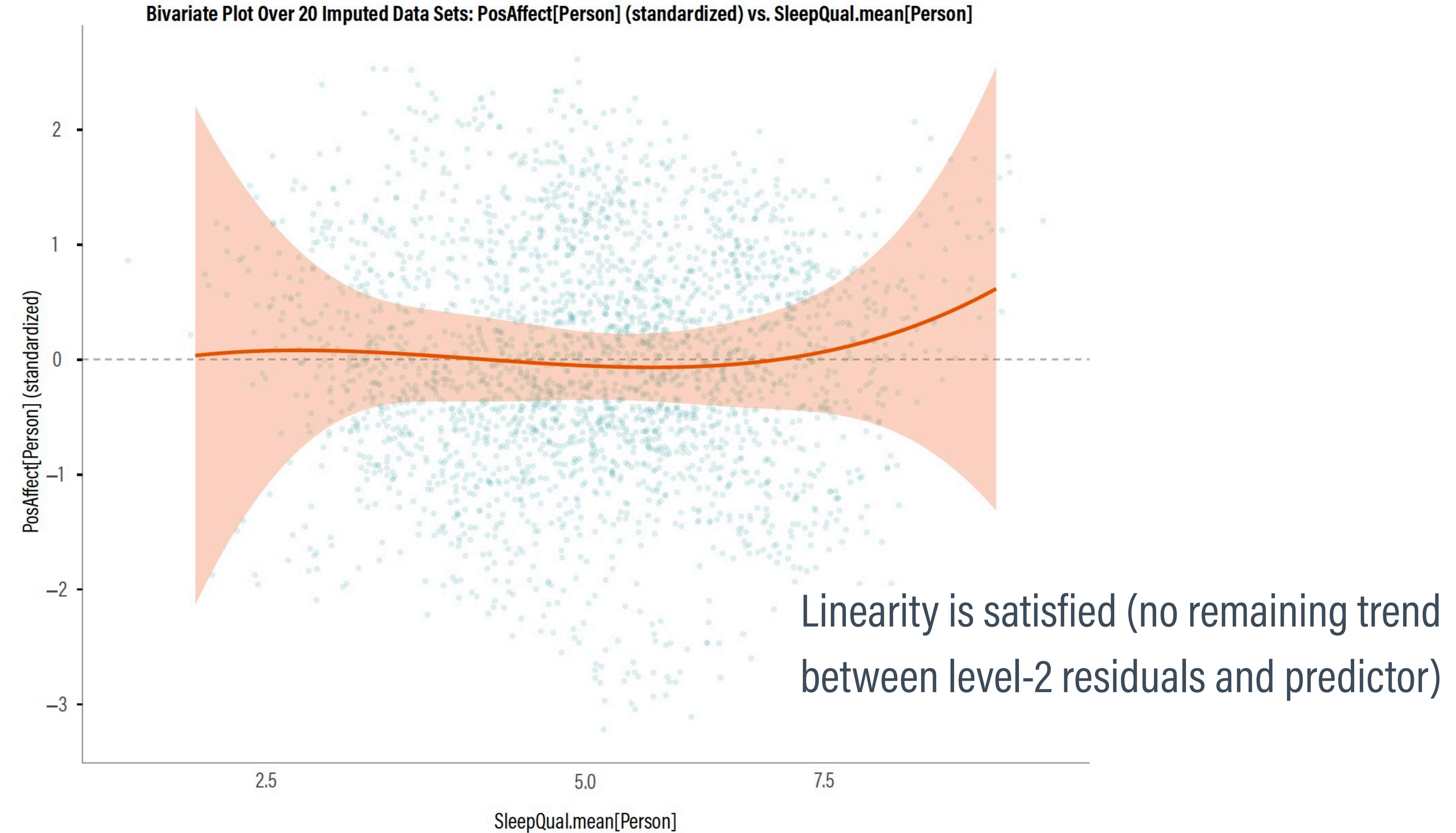
# LEVEL-1 RESIDUALS BY LEVEL-1 PREDICTOR

---



# LEVEL-2 RESIDUALS BY LEVEL-2 PREDICTOR

---

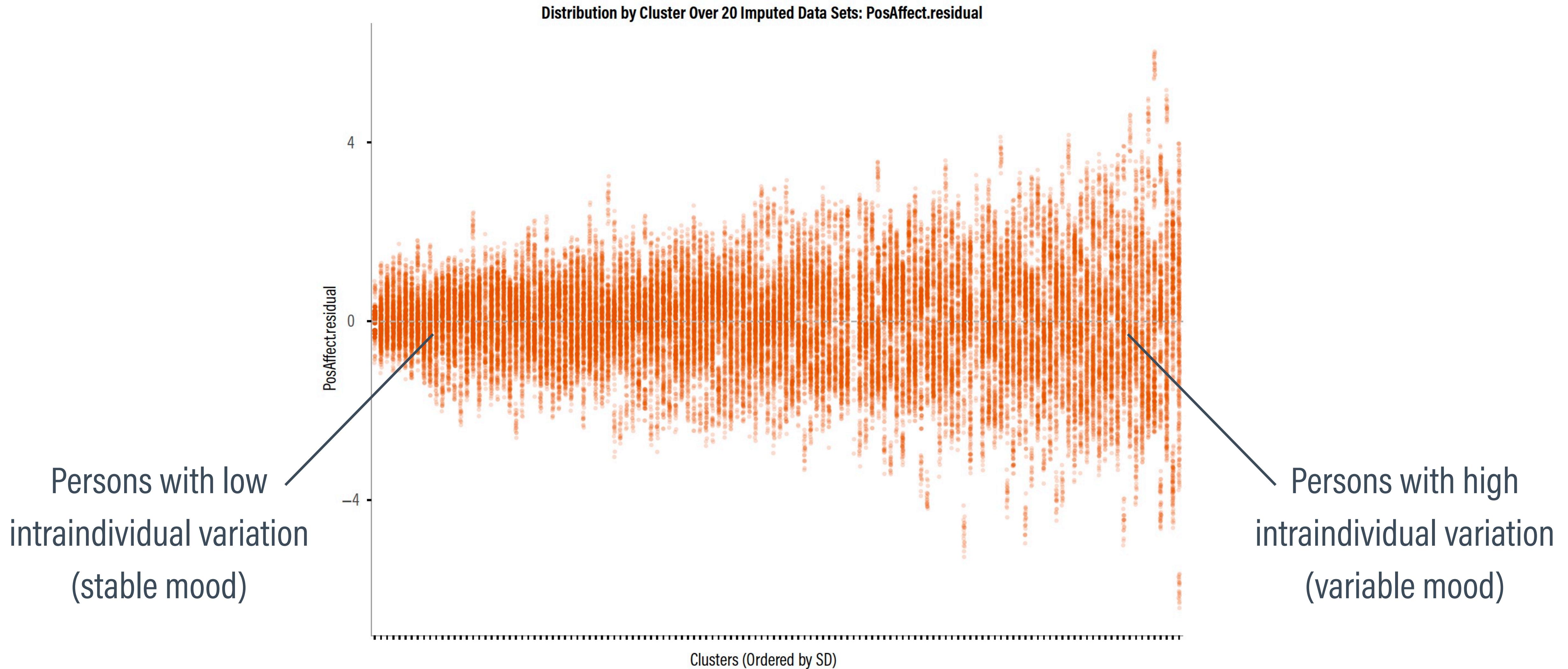


# VARIANCE HETEROGENEITY

---

- MLMs assume that the residual variance is constant (homogeneous) across clusters
- This assumption is often at odds with intensive data, where intraindividual variability differs across people (some have stable mood, others have large day-to-day fluctuations)
- Nonnormal (positively kurtotic) residuals are often a symptom of variance heterogeneity

# LEVEL-1 RESIDUALS BY CLUSTER



# SENSITIVITY ANALYSIS

---

- Blimp fits MLMs with cluster-specific variation
- Treat this as a sensitivity analysis: fit a model with cluster-specific residual variation (person-specific intraindividual variability) and evaluate whether the results change
- Are the main conclusions affected by assumptions about variance heterogeneity?

# BLIMP STUDIO SCRIPT 3.2

---

**DATA:** PainDiary.dat;

**VARIABLES:** Person Day PosAffect NegAffect Pain WorkGoal LifeGoal SleepQual Female Education  
Employment MarStatus NumDiagnose ActivityLevel PainAccept Catastrophize Stress Anxiety;

**CLUSTERID:** Person;

**CENTER:**

grandmean = SleepQual.mean; # defines the fixed (average) intercept as the grand mean

groupmean = SleepQual; # cwc with level-2 latent group means

**MODEL:** PosAffect ~ intercept SleepQual SleepQual.mean | intercept; # .mean invokes latent means

**BURN:** 10000;

**ITERATIONS:** 10000;

**SEED:** 90291;

**OPTIONS:** hev;

# RBLIMP SCRIPT 3 (MODEL 3)

---

```
model3 <- rblimp(  
  data = PainDiary,  
  clusterid = 'Person',  
  center = 'grandmean = SleepQual.mean; groupmean = SleepQual',  
  model = 'PosAffect ~ intercept SleepQual SleepQual.mean | intercept',  
  seed = 90291,  
  burn = 10000,  
  iter = 10000,  
  options = 'hev')  
output(model3)
```

# BLIMP OUTPUT

 = level-2 estimate

 = level-1 estimate

Outcome Variable: PosAffect

Group Mean Centered: SleepQual

Parameters	Estimate	StdDev	2.5%	97.5%	ChiSq	PValue	N_Eff
<hr/>							
Variances:							
L2 : Var(Intercept)	<b>1.845</b>	0.251	1.438	2.418	---	---	4565.552
Heterogeneity Index	0.241	0.042	0.174	0.336	---	---	3676.792
Q25% Residual Var.	<b>0.656</b>	0.036	0.587	0.727	---	---	6552.527
Q50% Residual Var.	<b>1.066</b>	0.055	0.963	1.180	---	---	8295.850
Mean Residual Var.	<b>1.337</b>	0.050	1.249	1.443	---	---	6775.414
Q75% Residual Var.	<b>1.747</b>	0.096	1.573	1.955	---	---	7606.960
<hr/>							
Coefficients:							
Intercept	<b>5.051</b>	0.150	4.760	5.352	1140.408	0.000	69.000
SleepQual	<b>0.167</b>	0.013	0.143	0.191	178.266	0.000	6998.401
SleepQual.mean[Person]	<b>0.622</b>	0.094	0.435	0.799	43.951	0.000	94.442
<hr/>							
...							
<hr/>							
Proportion Variance Explained							
by Coefficients	0.200	0.044	0.115	0.287	---	---	101.893
by Level-2 Random Intercepts	0.463	0.042	0.387	0.550	---	---	242.871
by Level-1 Residual Variation	0.211	0.023	0.170	0.258	---	---	430.938

# MODEL COMPARISON

---

- The stability of estimates between models suggests that the results were not materially affected by variance heterogeneity

Parameter	MCMC (Homogeneous)			MCMC (Heterogeneous)		
	Est.	Std. Dev.	p	Est.	Std. Dev.	p
Fixed intercept	5.00	0.13	< .001	5.05	0.13	< .001
Sleep (within-person)	0.17	0.01	< .001	0.17	0.01	< .001
Sleep (between-person)	0.60	0.09	< .001	0.62	0.09	< .001
Random intercept variance	1.84	0.25	—	1.85	0.25	—
Residual within-person variance	1.31	0.04	—	1.07	0.06	—