Bayesian Approaches to Statistics and Modeling Case Study • Part III

Mark Kurzeja

The Data

- Last time: we found that a normal regression model failed to capture skew and did not take into account the implicit groupings in the data
- Should we include the high school variable? Should we cluster based on IQ, age, or both? What belief are we expressing if we do?

	kid_score	mom_hs	mom_iq	mom_age
1	65	1	121	27
2	98	1	89	25
3	85	1	115	27
4	83	1	99	25
5	115	1	93	27
6	98	0	108	18

The Data

- Clustering Variables:
  - We could cluster on whether a mother has a certain range of IQs
  - We could cluster based on the mother's age
  - Whether the mother went to high school or not is another great clustering variable
- For simplicity, we are going to cluster based on whether a mother went to high school and their IQ

	kid_score	mom_hs	mom_iq	mom_age
1	65	1	121	27
2	98	1	89	25
3	85	1	115	27
4	83	1	99	25
5	115	1	93	27
6	98	0	108	18

- We will use
  - Low IQ = $IQ < 85$
  - Medium IQ = $85 \leq IQ \leq 115$
  - High IQ = $115 < IQ$

Ra… hi… Belief about the World → Collect Data → Bayesian Update

The Model

- We are going to allow all three parameters to vary according to whether a mother attended HS and their IQ group
  - For example, we will have a different β_0, β_1, and β_2 for mothers who are high IQ and attended HS, for those mothers that did not attend HS but also have a high IQ, etc.
- This model now has six sets of $\beta_0's$, $\beta_1's$ and $\beta_2's$ 2x3
- We are also going to say that each of the $\beta_0's$ come from a common distribution. We will also do the same with the $\beta_1's$ and the $\beta_2's$

Belief about the World → Collect Data → Bayesian Update

The Data

- Modeling in a Bayesian way forces us to state a lot of our assumptions outright
- Stating all assumptions allows someone to challenge them
- We will used a hierarchical model with varying intercepts to capture this belief
- Is

Key Point: Always ask **why** a modeling decision is being made

The belief is that the relationship between the IQ of the mother and the child varies based on the mothers education and the mothers IQ group

The Data

- Is this belief "right"? *Of course not*
- All models are subjective, Bayesian or not
- It is important to remember this as we increase the complexity of the model – these are modeling choices based on updated beliefs and not ground truth

Key Point: *All* modeling is subjective

Belief about the World → Collect Data → Bayesian Update

The Model

Multilevel Regression Form

Response Variable

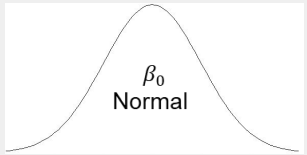
Slope on IQ for a given high school status and IQ group

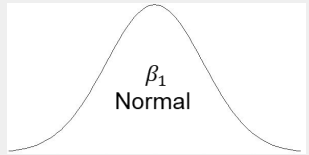
$$ChildIQ = \beta_{0,HS,IQG} + \beta_{1,HS,IQG} \times (momIQ) + \beta_{2,HS,IQG} \times (momAge)$$

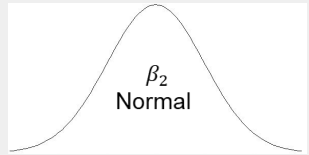
Intercept term for a given high school status and IQ group

Slope on age for a given high school status and IQ group

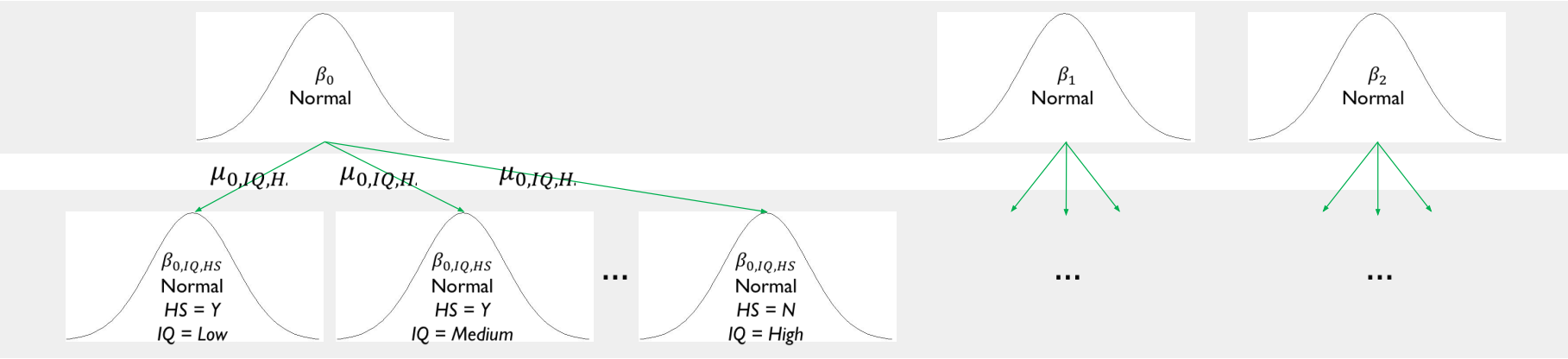
Belief about the World → Collect Data → Bayesian Update



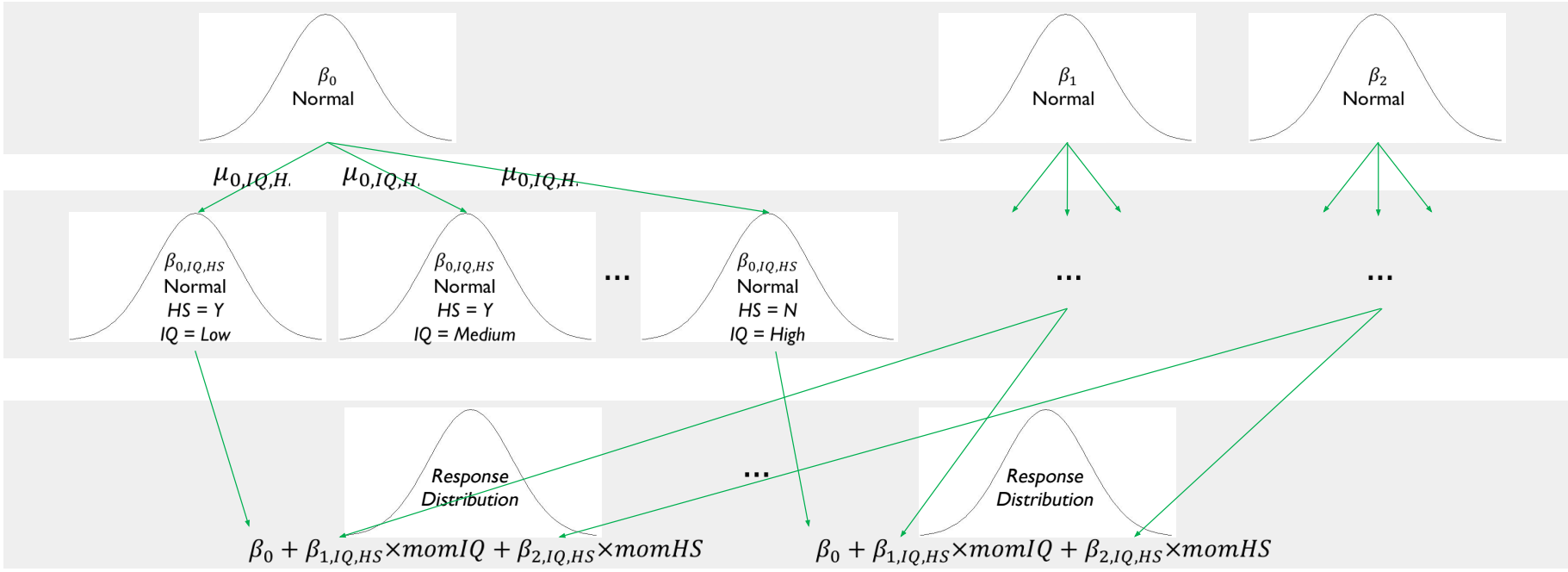




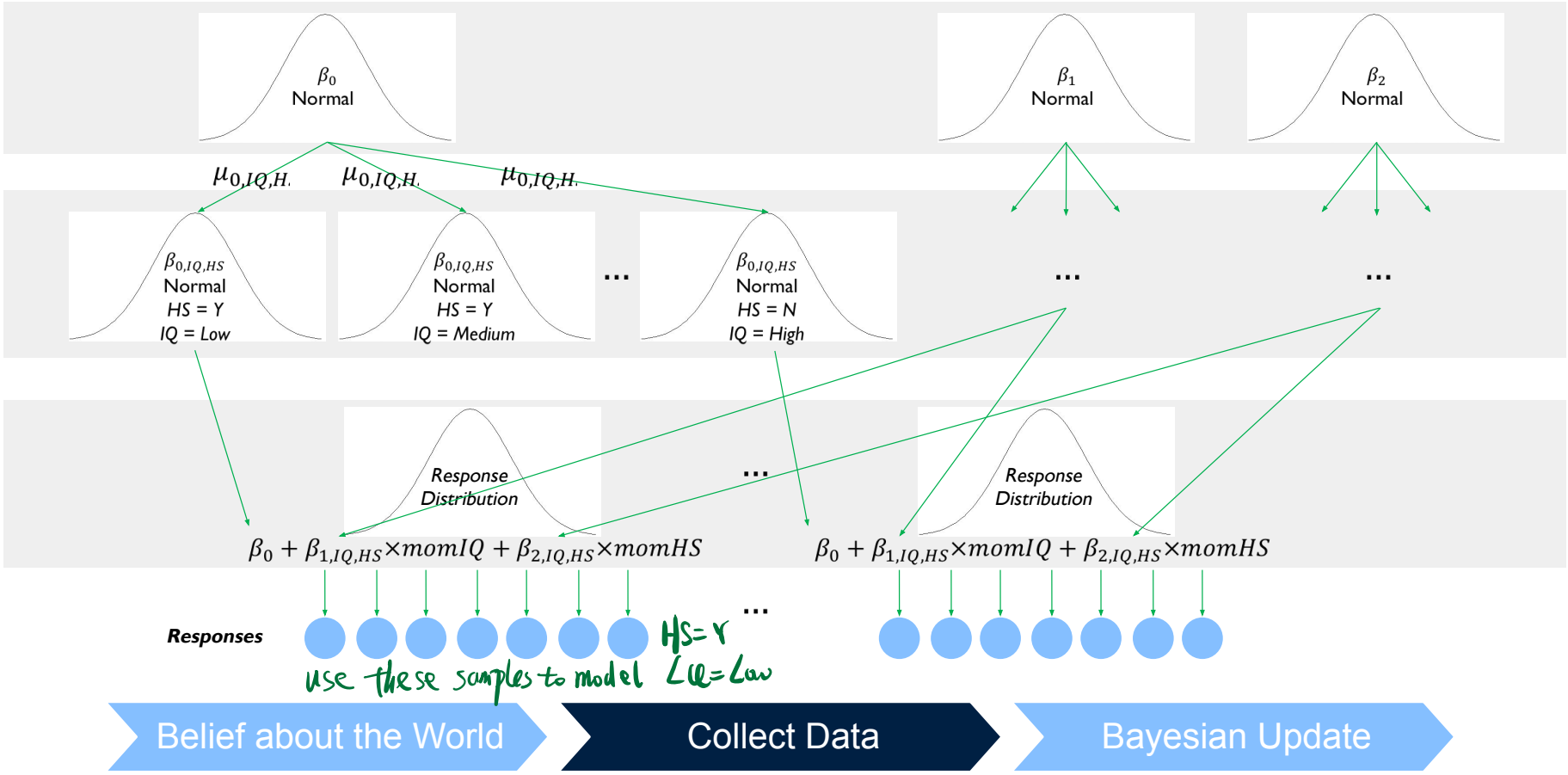
Belief about the World → Collect Data → Bayesian Update



Belief about the World → Collect Data → Bayesian Update



Belief about the World → Collect Data → Bayesian Update


β_0 Normal
β_1 Normal
β_2 Normal
$\mu_{0,IQ,H}$
$\mu_{0,IQ,H}$
$\mu_{0,IQ,H}$
$\beta_{0,IQ,HS}$ Normal HS = Y IQ = Low
$\beta_{0,IQ,HS}$ Normal HS = Y IQ = Medium
$\beta_{0,IQ,HS}$ Normal HS = N IQ = High
Response Distribution
Response Distribution
$\beta_0 + \beta_{1,IQ,HS} \times momIQ + \beta_{2,IQ,HS} \times momHS$
$\beta_0 + \beta_{1,IQ,HS} \times momIQ + \beta_{2,IQ,HS} \times momHS$
Responses
HS=Y
use these samples to model IQ=Low
Belief about the World
Collect Data
Bayesian Update

The Model

- Hierarchical models often have a lot of moving parts (as we can see in the graphic)
- They can model complicated relationships in a more intuitive way than some other methods of modeling
- Downfalls: fitting one of these models in a Bayesian framework can be computationally difficult

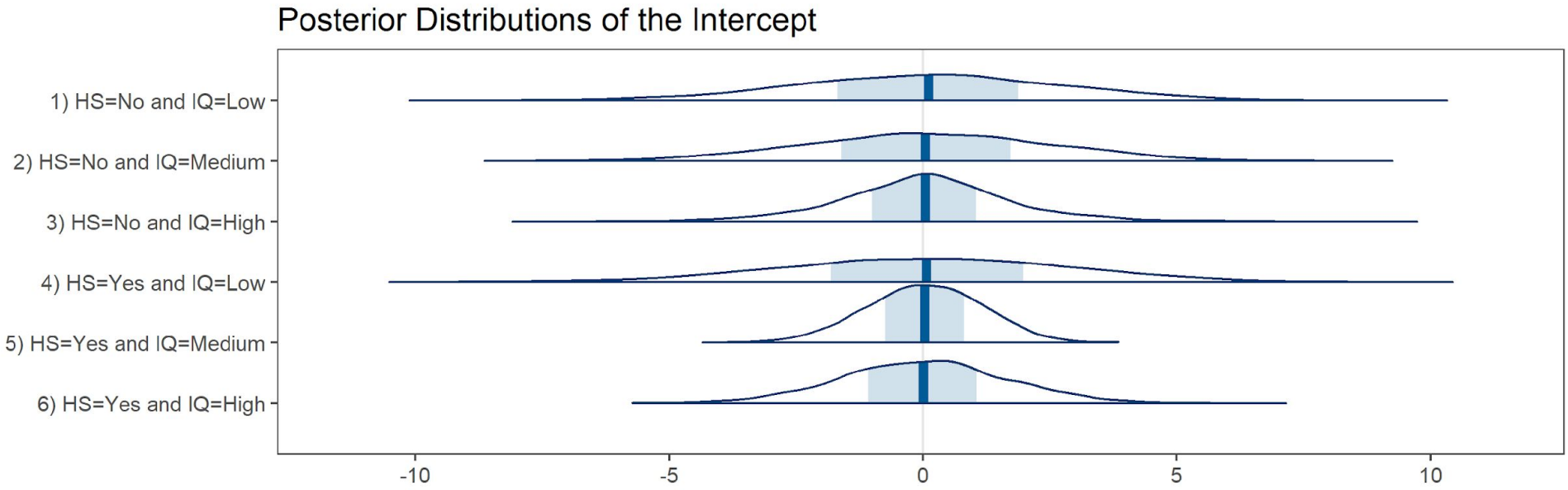
The Model - Updates

- Include six new intercepts ($\beta_0's$), $\beta_1's$, $\beta_2's$, and $\sigma_i's$ – each constrained by a hierarchical relationship
- Change the distribution of the errors from a normal to a skewed normal to account for the left-skew that we observed in the posterior predictive fits

The Model

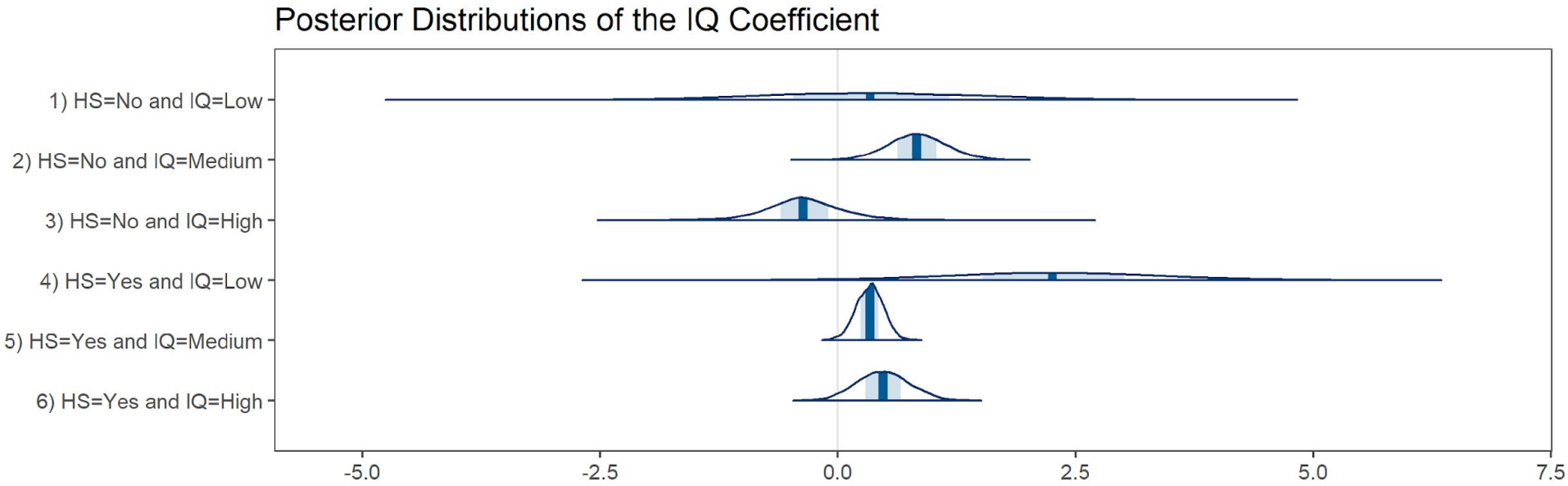
- Just like before, when we run the model we will get posterior distributions on each of the parameters
- This time, we can compare each subgroup's intercept and slope to see if they have any relationships
- Why not incorporate a non-linear term (like IQ^2) instead?
  - Expressing belief that different IQ groups come from different distributions and not a non-linear extension of a more global process

The Model – The Posteriors



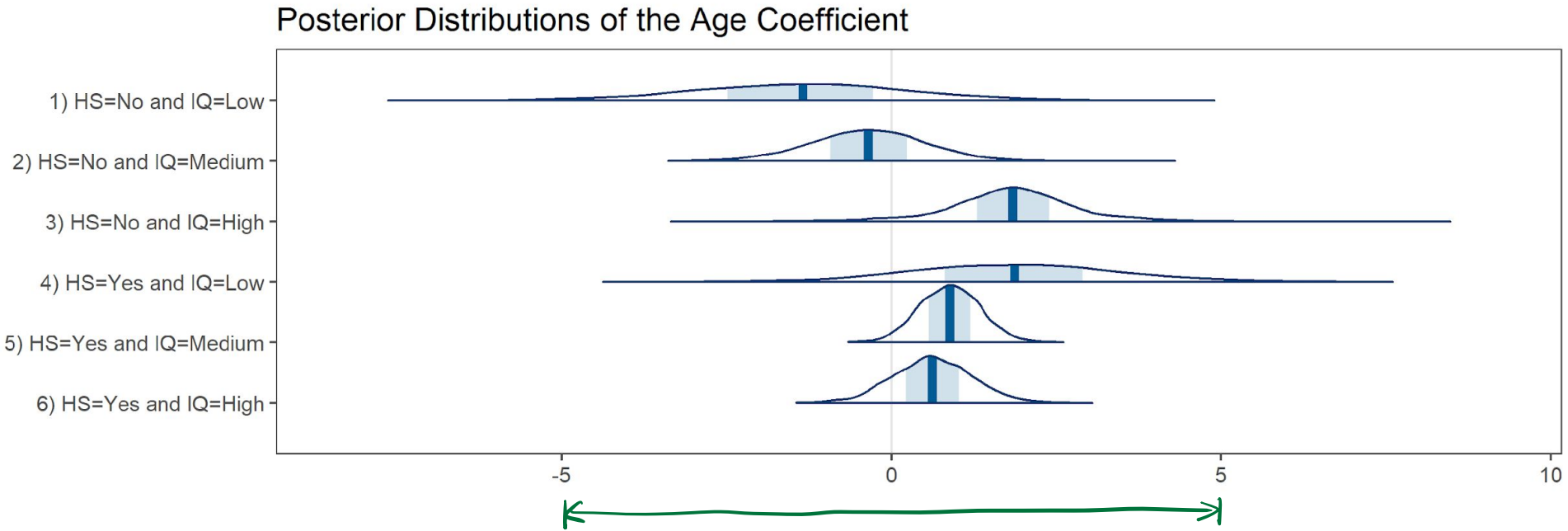
Belief about the World → Collect Data → Bayesian Update

The Model – The Posteriors

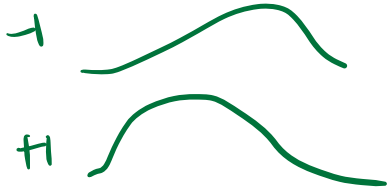


Belief about the World → Collect Data → Bayesian Update

The Model – The Posteriors



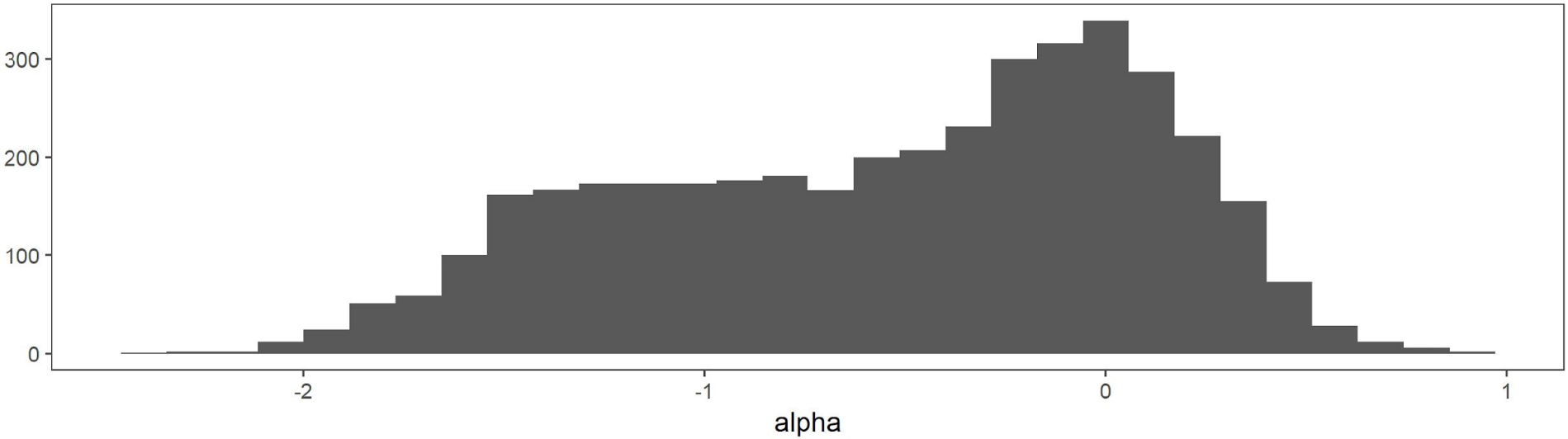
Belief about the World → Collect Data → Bayesian Update



The Model – The Posteriors

Posterior Distribution of Skew Parameter

The skew parameter is predominatly negative indicating that the distribution of errors is left-skewed

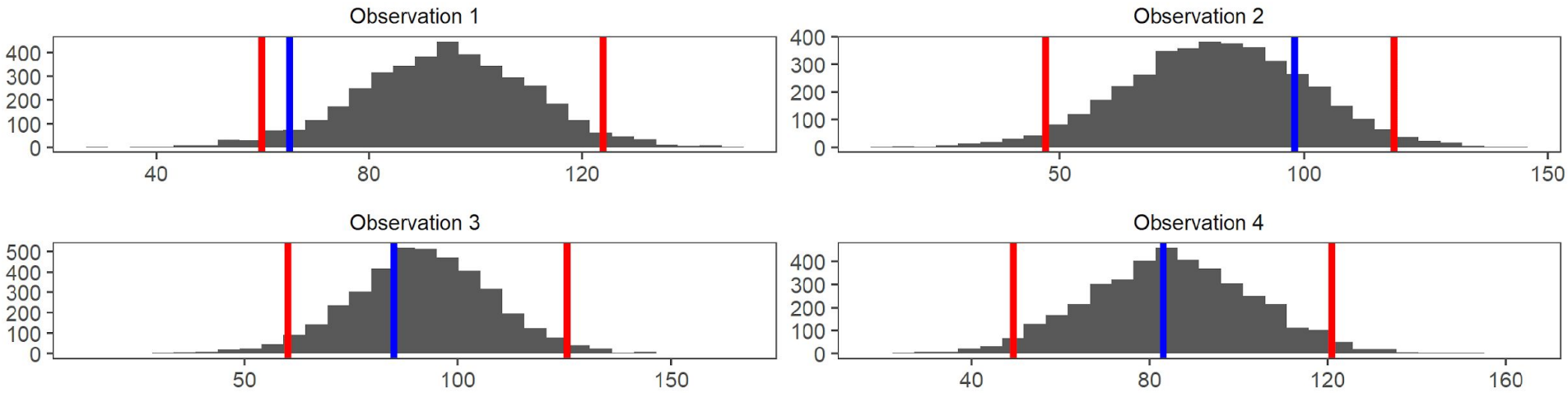


Belief about the World → Collect Data → Bayesian Update

The Model – The Posteriors

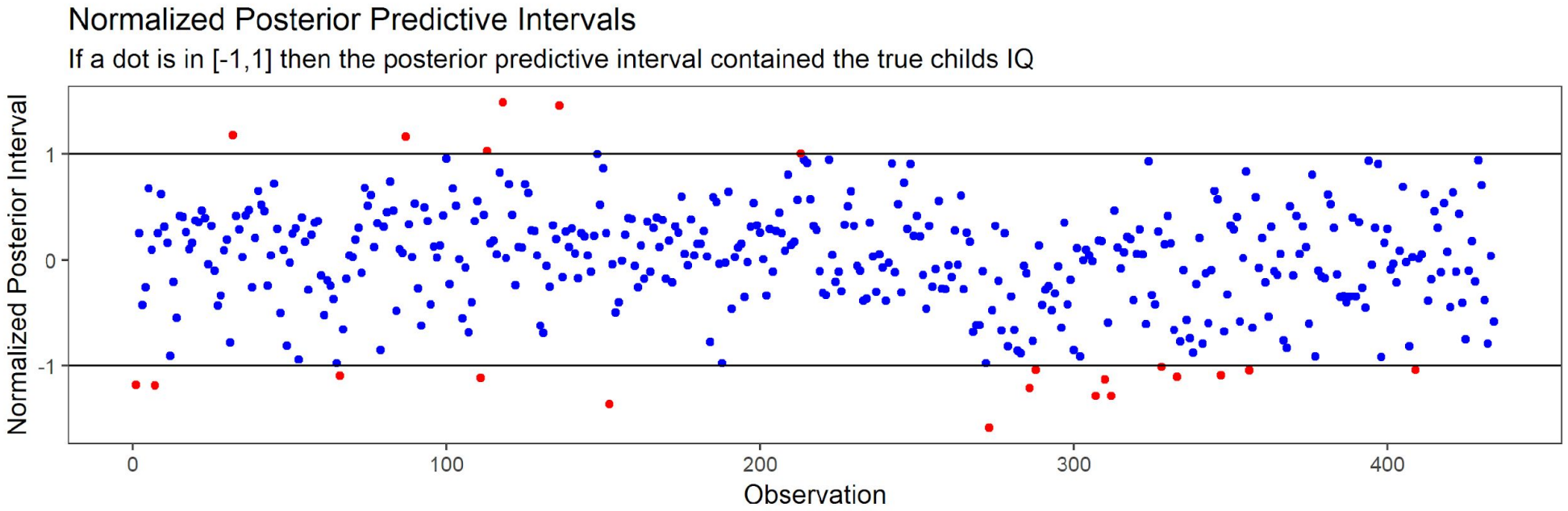
Posterior Predictive Intervals for First Four Observations

The blue line is the observed value for Childs IQ. The red lines are the 95% predictive interval



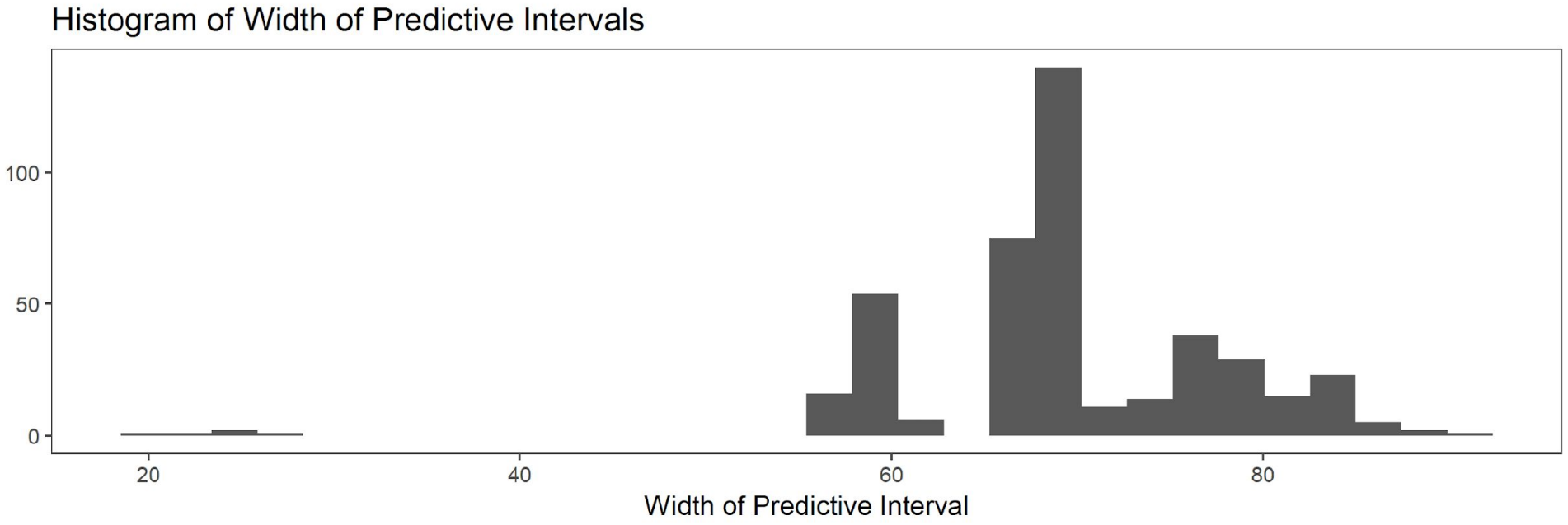
Belief about the World → Collect Data → Bayesian Update

The Model – The Posteriors



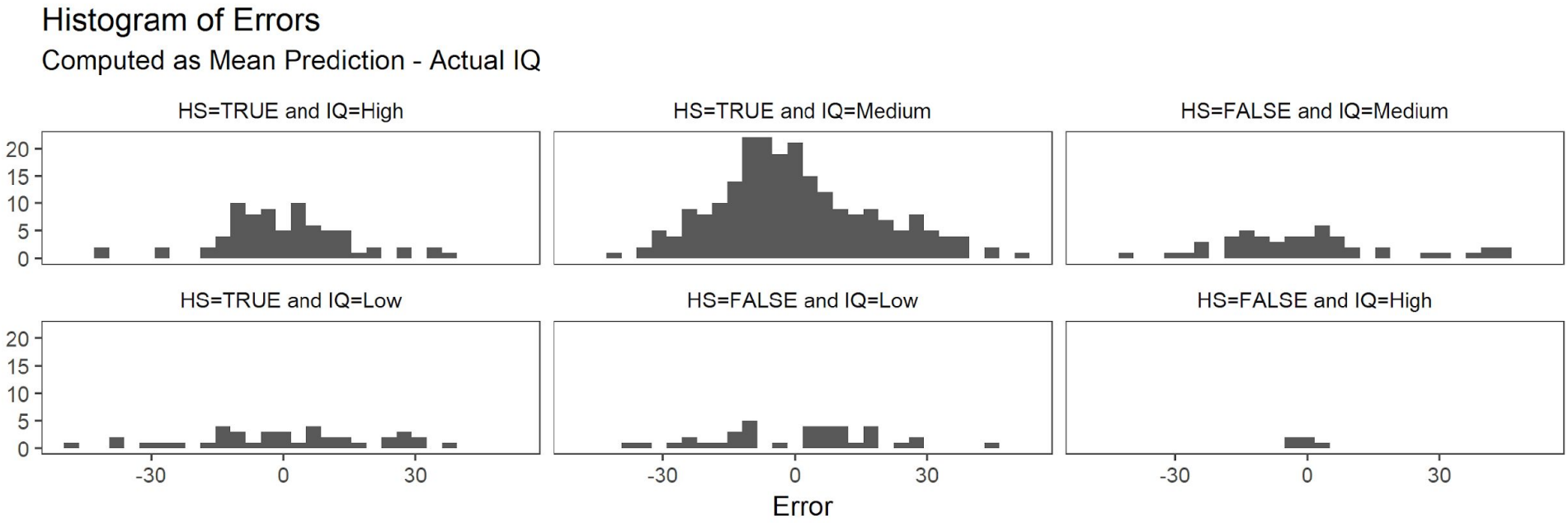
Belief about the World → Collect Data → Bayesian Update

The Model – The Posteriors



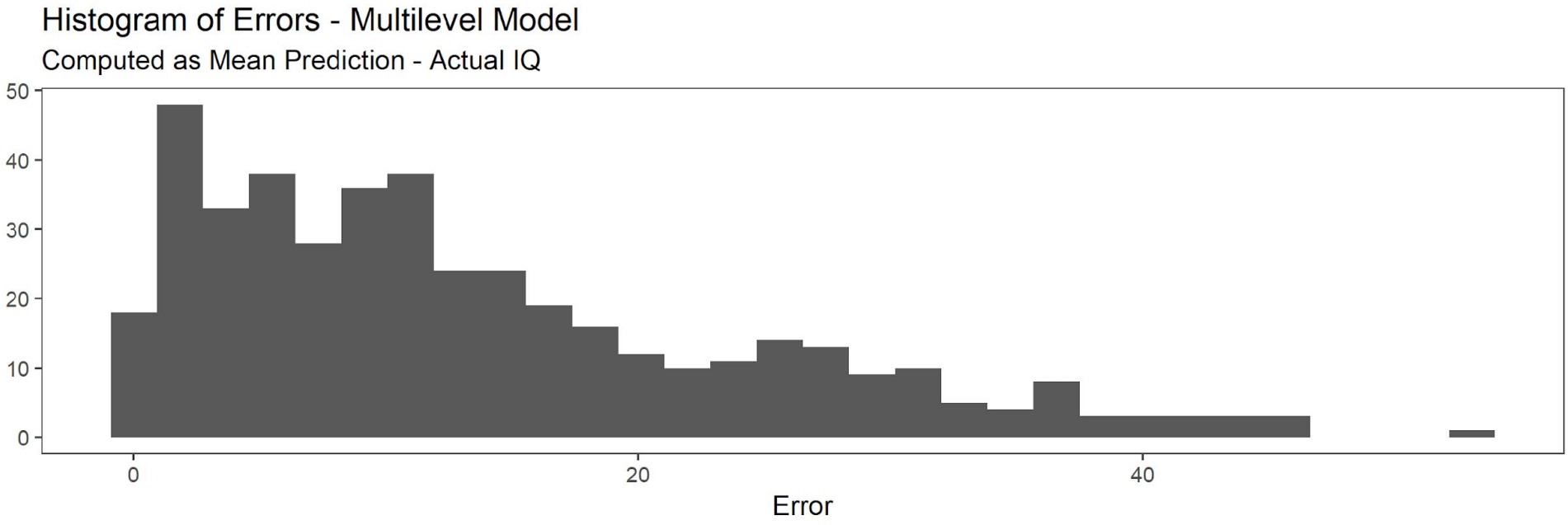
Belief about the World → Collect Data → Bayesian Update

The Model – The Posteriors



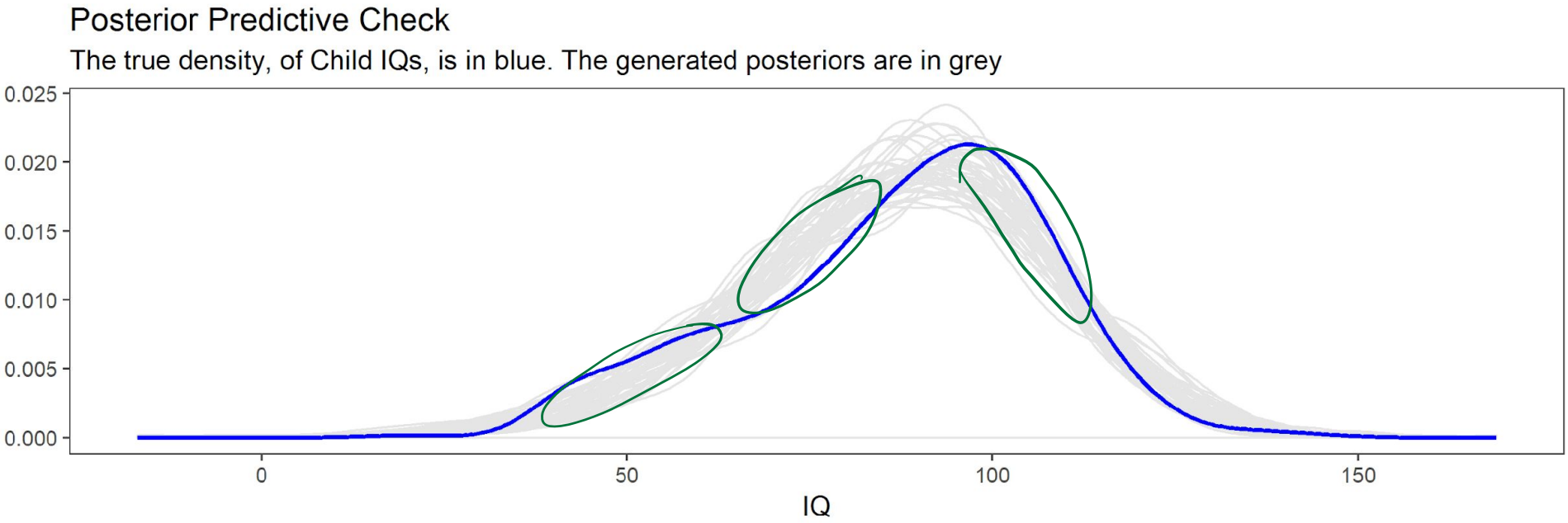
Belief about the World → Collect Data → Bayesian Update

The Model – The Posteriors



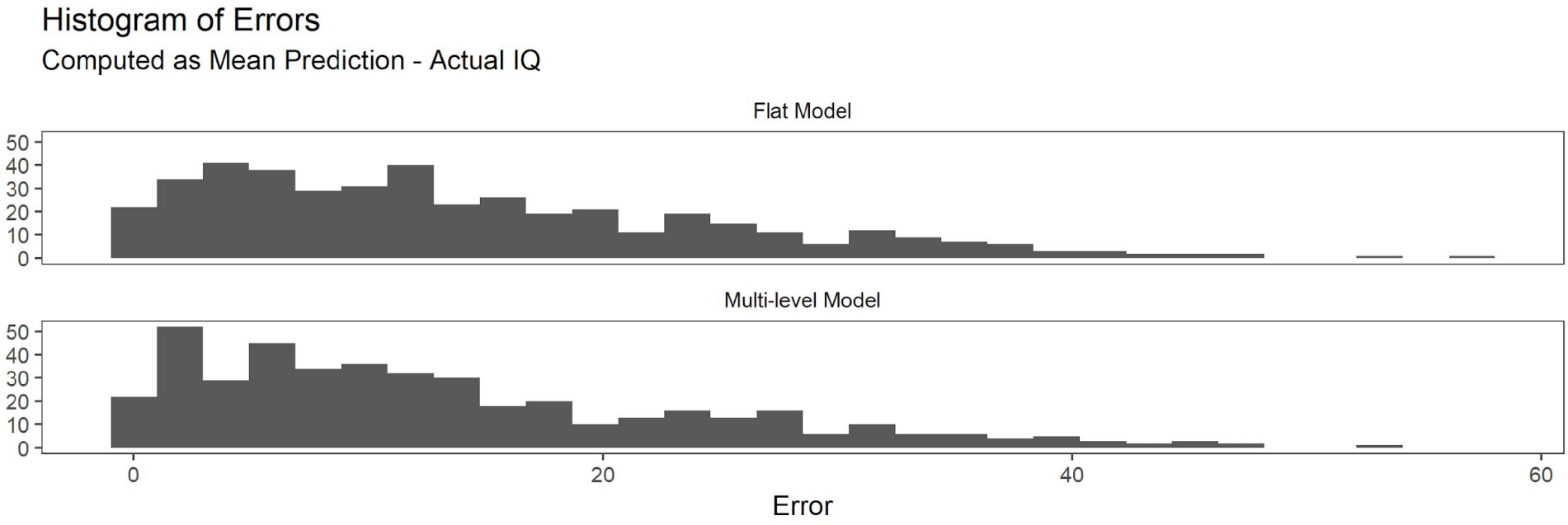
Belief about the World → Collect Data → Bayesian Update

The Model – The Posteriors



Belief about the World → Collect Data → Bayesian Update

The Model – The Posteriors



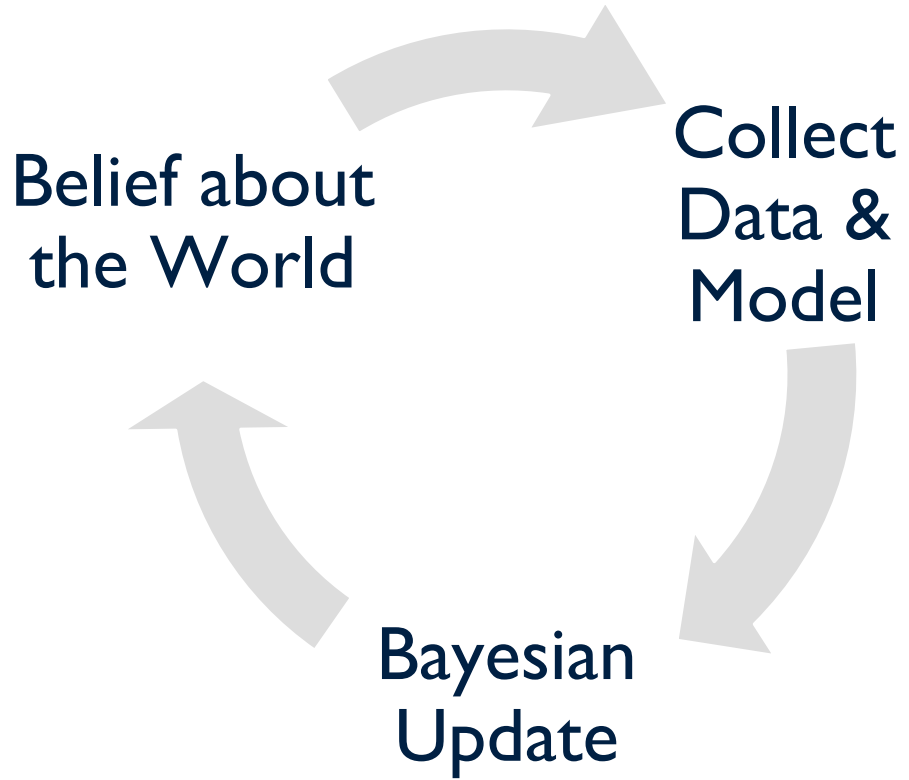
Belief about the World → Collect Data → Bayesian Update

The Model - Observations

- The multi-level model, for this data, isn't a massive improvement in prediction ability but is an improvement in inference – we can now tease out the individual group effects
- It provides insight into the differences between the slopes of the parameters
  - Increased variance in parameter estimates as a result
- Next steps in modeling:
  - Account for the over-inflated left-tail using a mixture distribution or another technique
  - 

Parting Notes

- Bayesian models are computationally intensive but can model very complex relationships, perform regularization, incorporate previous analysis, and express uncertainty intuitively within the same model
- Methods are becoming more feasible due to advances in sampling techniques and computational power



Parting Notes

- Modeling is an iterative process that is filled with assumptions and is subjective
- Capturing the variance of an estimate is sometimes more important than the estimate itself
- Large posterior intervals → Large uncertainty in estimate given the current model and data!