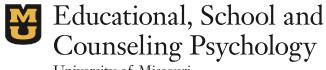
An Introduction to Missing Data and Bayesian Statistics with Blimp

Brian T. Keller

Assistant Professor
Statistics, Measurement, & Evaluation in Education
Educational, School, & Counseling Psychology
University of Missouri

Developer, Blimp



Workshop Content

Blimp available at

`https://www.appliedmissingdata.com/blimp`

Workshop content at

`https://github.com/blimp-stats/Yonsei-Workshop`

Acknowledgements

Craig Enders, Ph.D. (UCLA)

Han Du, Ph.D. (UCLA)

Jiwon Kim, Ph.D. (UT Austin)

Suyoung Kim, M.A. (UT Austin)

IES Institute of Education Sciences

Workshop Outline

Day 1:

- An Introduction to Missing Data and Bayesian Statistics with Blimp

Day 2:

- Finish remaining material from Day 1.
- Latent Variable Models with Blimp

Day 1's Overview

- Missing Data Processes
- Introduction to Bayesian Statistics
- Fitting Regression Models in Blimp
- Understanding Blimp Output
- Incomplete Categorical Variables
- Interaction Effects in Blimp

Why Bayesian statistics?

Things Bayesian approaches are good at...

- Direct estimation for complex models with missing data
- Introduction of mixed response metrics (normal, binary, ordinal, etc.)
- Nonlinear effects (interactions, curvilinear, etc.)
- Multilevel data (random effects, interactions, heterogenous variance)
- Complex latent variable modeling (interactions, multilevel)

Missing Data Processes

Introduction to Missing Data

- Generally, statistical methods are developed to make population inferences from ***rectangular data sets***, where the rows are observations and the columns represent random variables.

Data

Y	X
4	2
1	7
1	5
1	8
9	2
3	3
4	8
2	1

- Missing or incomplete data is concerned with instances where some entries in our data set are not observed.
- Missing data can occur on different types of variables in the design: covariates, mediators, moderators, final outcome variables of interest.

Data

Y	X
4	2
NA	7
1	5
1	NA
9	2
NA	3
4	NA
NA	1

Partitioning the Data

Complete		=	Observed		+	Missing		Indicators	
Y	X		Y	X		Y	X	M_Y	M_X
4	2		4	2				0	0
1	7		**NA**	7		1		1	0
1	5		1	5				0	0
1	8		1	**NA**			8	0	1
9	2		9	2				0	0
3	3		**NA**	3		3		1	0
4	8		4	**NA**			8	0	1
2	1		**NA**	1		2		1	0

Two Distinct Concepts

- The missingness pattern describes the configuration of the observed and missing values in a data set.

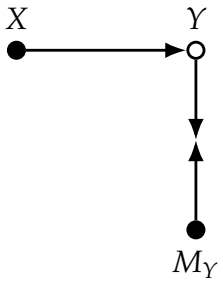
- The missingness mechanism describes the relationship between the missingness and the values of variables in the data.

Missing Data Mechanisms

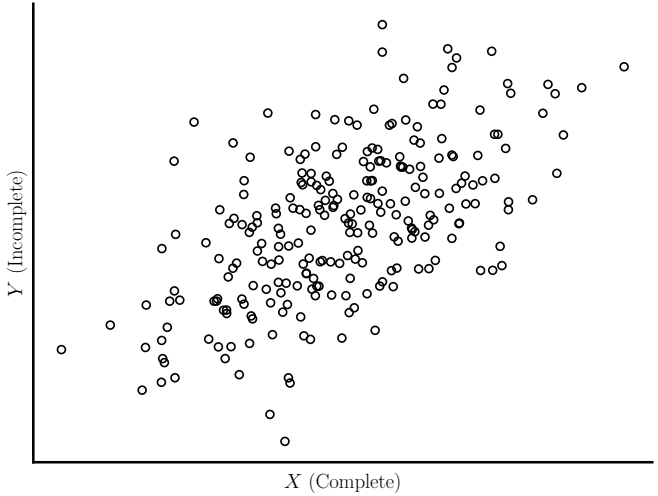
- Conceptually, we are treating the missing data indicator as a binary variable representing if an observation for a variable is missing ($M = 1$) or observed ($M = 0$).
- The missing data mechanism is a classification of the relationship between the missing data indicator to the variables in the data set.
- Missingness may be independent of the data, or it could relate to the observed or missing parts (or both).

Missing Completely at Random (MCAR)

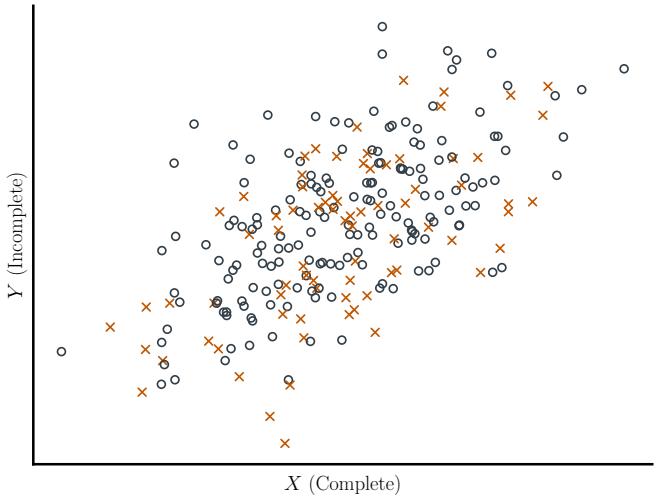
Missing completely at random stipulates that the probability of observations being missing is ***unrelated*** — be it directly or via spurious relationships — to the data.



Scatterplot of X and Y

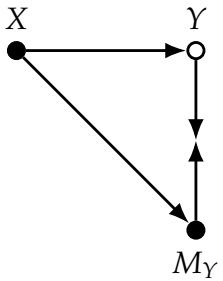


MCAR – Scatterplot of Missing and Observed Values

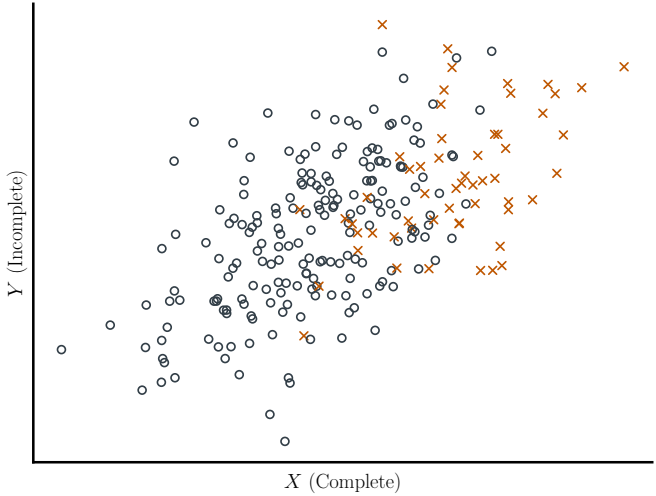


Missing at Random (MAR)

Missing at random stipulates that the probability of observations being missing is related — be it directly or via spurious relationships — to the ***observed data but not the missing data***.



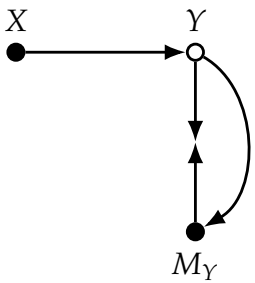
MAR – Scatterplot of Missing and Observed Values



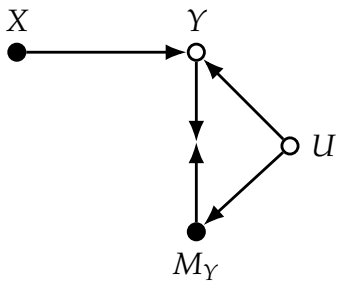
Missing Not at Random (MNAR)

Missing not at random stipulates that the probability of observations being missing is related — be it directly or via spurious relationships — to the ***missing data***.

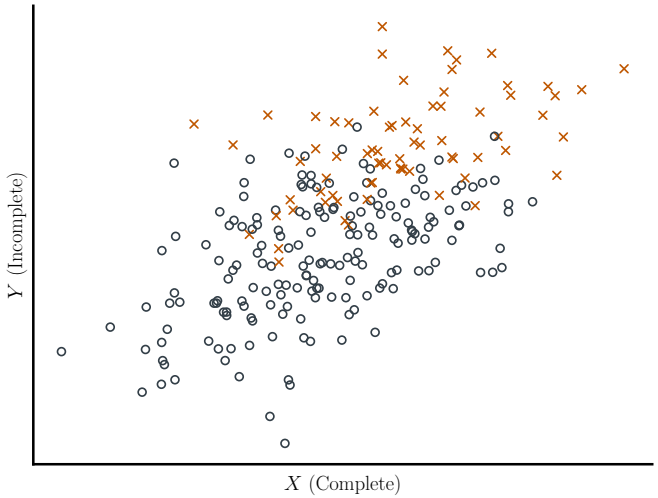
Direct MNAR



Indirect MNAR



MNAR – Scatterplot of Missing and Observed Values



Mechanisms Summary...

- Missing completely at random $\Rightarrow$ haphazard missingness
- Missing at random $\Rightarrow$ systematic missingness solely related to observed values
- Missing not at random $\Rightarrow$ systematic missingness related to the value of the missing observation

Working Assumptions

- Bayesian methods and modeling approach we discuss assume that the missing process is MAR.
- Missing at random is untestable because it involves knowing the "Truth" about the unobserved scores.
- These methods can extended to MNAR with additional modeling (but not discussed in this workshop).

Introduction to Bayesian Statistics

Why Bayesian statistics?

Bayesian[1] v. Frequentist[2]

[1] What I am.

[2] If you don't know, probably what you are.

Frequentist

- Probability is defined by an event's hypothetical relative long run frequency.
- Probability is applied to hypothetical future sampled data (e.g., sampling distributions for estimates).
- Parameters are 'fixed' at some unknown value in the population.

Bayesian

- Probability is defined by one's degree of belief
- Probability can be applied to anything really.
- Parameters are 'random' and represented by a probability distribution of plausible values. We try to estimate the distribution based on one's belief.

Frequentist Approach

- A Frequentist point of view finds the single fixed value that is most likely given the sample we observed (via maximum likelihood, FIML).

- FIML uses the probability of the data given the parameter values—the likelihood of the data given a fixed parameter in the population; $P\left(\text{data} \mid \text{parameters}\right)$.

Bayesian Approach

- Bayesian takes this one step further and estimates the probability of the parameters given the data—$P\left(\text{parameters} \mid \text{data}\right)$.

- Because most parameters are continuous, this means we characterize all values that the parameter potentially could take via a probability distribution.

How do we obtain the probability of the parameters given the observed data when what we have is the reverse?

$$P\left(\text{parameters} \mid \text{data}\right) \neq f\left(\text{data} \mid \text{parameters}\right)$$

What we want to know

What we have

Bayes' Theorem

$$P\left(\text{parameters} \mid \text{data}\right) = \frac{P\left(\text{data} \mid \text{parameters}\right) P\left(\text{parameters}\right)}{P\left(\text{data}\right)}$$

$P\left(\text{parameters} \mid \text{data}\right)$: The posterior distribution. The probability of the parameters given the data (what we want).

Bayes' Theorem

$$P\left(\text{parameters} \mid \text{data}\right) = \frac{P\left(\text{data} \mid \text{parameters}\right) P\left(\text{parameters}\right)}{P\left(\text{data}\right)}$$

$P\left(\text{data} \mid \text{parameters}\right)$: The likelihood. The probability of the data given the parameters (what we observe).

$P\left(\text{parameters}\right)$: The prior distribution. The probability of the parameters overall. This represents the a priori knowledge we have about the parameters.

Bayes' Theorem

$$P\left(\text{parameters} \mid \text{data}\right) = \frac{P\left(\text{data} \mid \text{parameters}\right) P\left(\text{parameters}\right)}{P\left(\text{data}\right)}$$

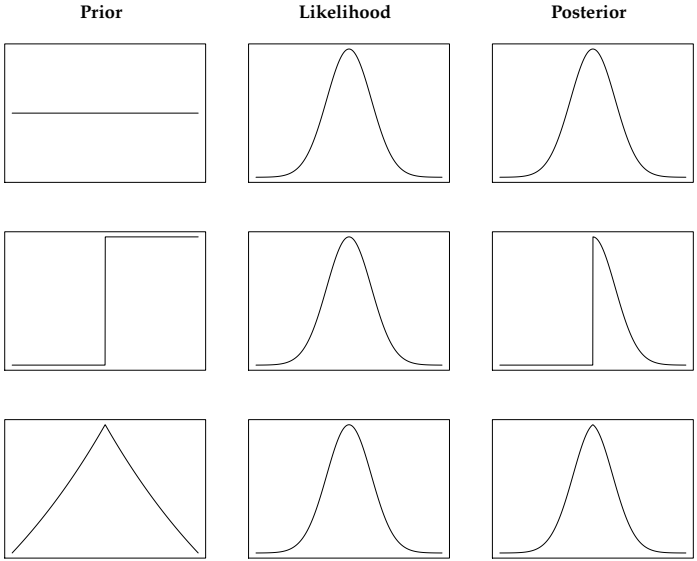
$P\left(\text{data}\right)$: A scaling term that makes the sum of $P\left(\text{parameters} \mid \text{data}\right)$ equal one. We are not interested in this.

Prior Distribution

- Bayesian analyses require prior distributions that encode our beliefs about the parameter values prior to analyzing the data.
- Blimp adopts non-informative (diffuse) priors by default that impart as little information as possible (i.e., let the data do the talking).
- A diffuse prior for means and coefficients conveys that all possible parameter values are equally likely (a flat distribution).

Posterior Distributions

- The prior and likelihood function as two data sources that merge to define the posterior distribution of the parameters.
- The posterior describes the distribution of plausible values that could have produced the particular data given our model.
- Instead of estimates varying around a fixed parameter (Frequentist), the parameters vary around a fixed data set (Bayesian).


Prior
Likelihood
Posterior

Bayesian Statistics and this Workshop

- We will largely ignore the philosophical debate regarding what probability ought to represent.
- Instead, we will focus on using it practically and in a framework that does not require you to be a Bayesian.
- The priors we will use are generally diffuse and provide little information. These will generally fall under what is called an 'uninformative' or a 'weakly informative' prior.

Workshop Data

Chronic Pain Data – Variable Definitions

Name	Definition	Missing %	Range
`disab`	Psychosocial disability composite	9.1	10 to 34
`depress`	Depression composite score	13.5	7 to 28
`severity`	Severe pain dummy code	7.3	0 or 1
`male`	Biological sex dummy Code	0.0	0 or 1

One Predictor Regression Illustration

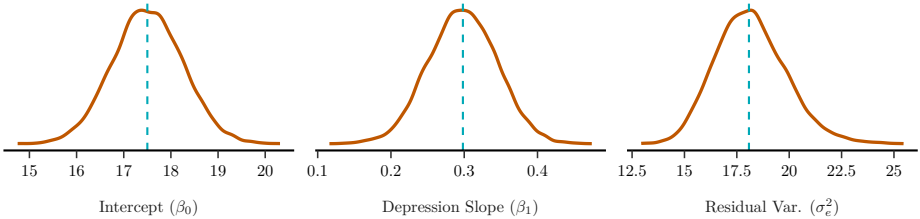
- A study that is investigating the relationship of psychological disability predicted by depression scores.

$$\texttt{disab}_i = \beta_0 + \beta_1 \left(\texttt{depress}_i\right) + e_i$$

Both measures have missing values that must be handled.

Posterior Distributions

	Median	SD	CI 2.5%	CI 97.5%
Intercept (β_0)	17.50	0.74	16.05	18.97
Depression Slope (β_1)	0.30	0.05	0.21	0.39
Residual Var. (σ^2_e)	18.08	1.69	15.17	21.79



Estimator Comparison

- Two approaches provide numerically equivalent results and inferences.

	Bayesian				Frequentist (FIML)			
	Median	SD	CI 2.5%	CI 97.5%	Estimate	SE	CI 2.5%	CI 97.5%
Intercept (β_0)	17.50	0.74	16.05	18.97	17.55	0.74	16.10	19.00
Depression Slope (β_1)	0.30	0.05	0.21	0.39	0.30	0.05	0.21	0.39
Residual Var. (σ^2_e)	18.08	1.69	15.17	21.79	17.49	1.68	14.19	20.79

Markov Chain Monte Carlo (MCMC)

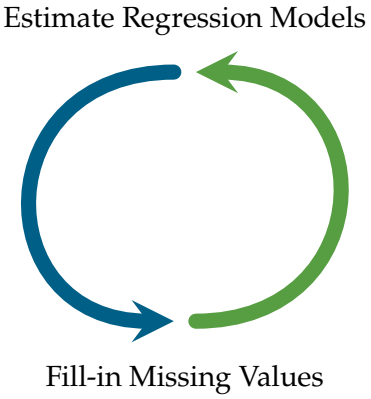
- A Bayesian analysis involves estimating the distribution of model parameters and the missing values by sampling methods.
- Typically, MCMC breaks the complex problem involving multiple unknown quantities (parameters and missing values) into separate steps.
- Each step estimates one unknown at a time, treating the current values of all other quantities as known constants.

MCMC Estimation

Do for $t = 1$ to T iterations:

- Estimate focal model parameters given filled-in data
- Estimate predictor model parameters given filled-in data
- Fill-in missing outcome scores given focal model parameters
- Fill-in missing predictor scores given focal and predictor model parameters

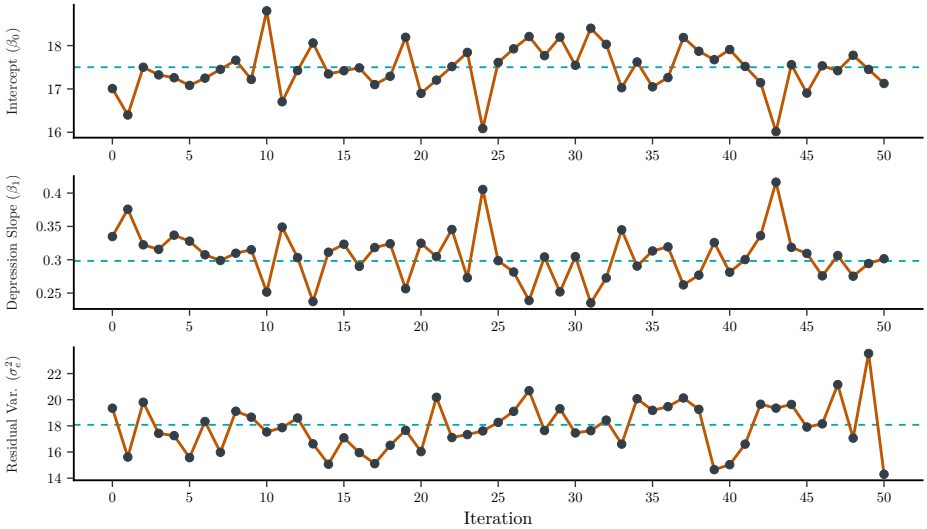
Repeat



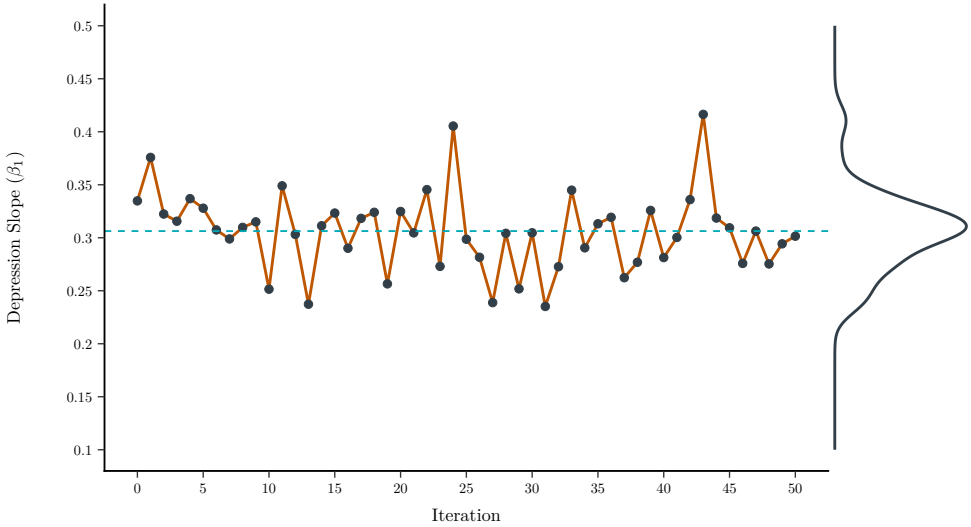
What do I mean by "Estimation"?

- A posterior distribution characterizes the probability of a parameter's "true" value given the data we observed and the model.
- MCMC uses computer simulation (random number generation) to sample from this distribution. We can then use descriptive statistics to summarize the posterior distribution.
- Parameters and missing data scores are unknown, so their values constantly change with new samples during the MCMC sequence.

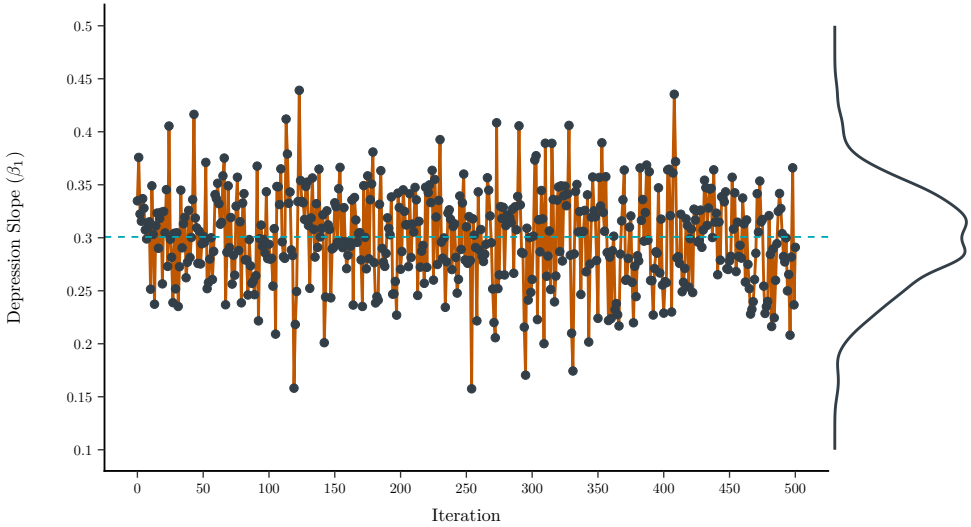
Parameters from 50 MCMC Iterations



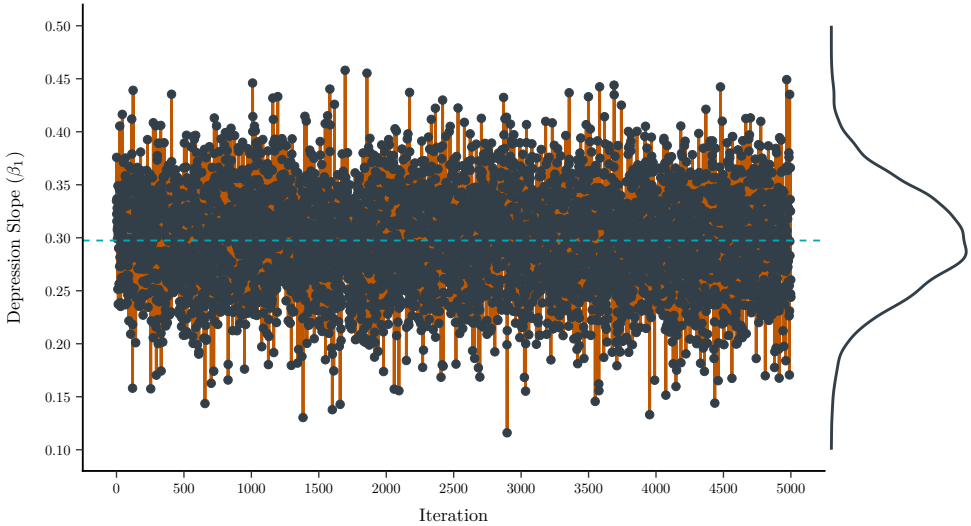
Depression Slope from 50 MCMC Iterations



Depression Slope from 500 MCMC Iterations



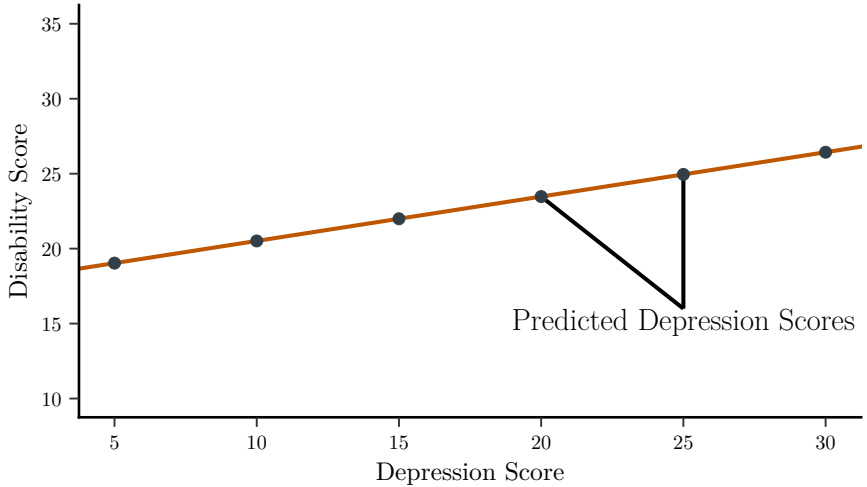
Depression Slope from 5000 MCMC Iterations



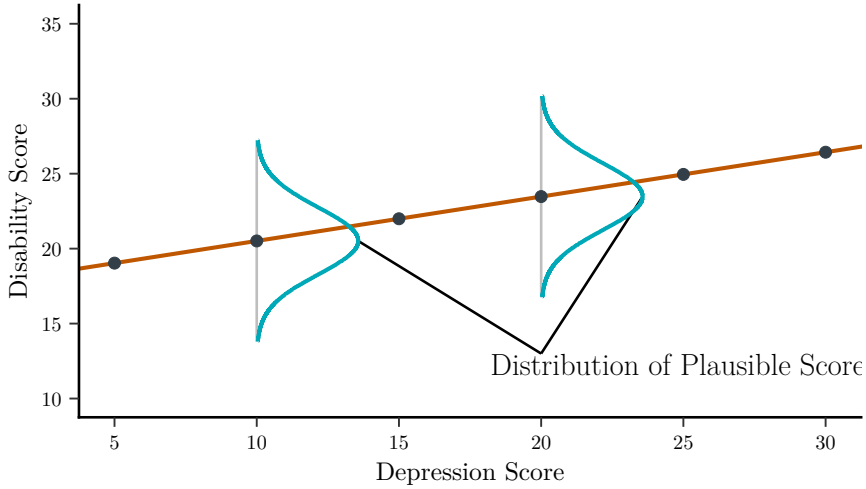
Filling-in Missing Data (Imputation)

- Missing scores are filled-in or "imputed" by randomly drawing replacement scores from a distribution of plausible values.
- Like parameters, we draw a new set of imputed scores for each missing variable in each iteration.
- Imputing predictors is more complex than imputing outcomes.

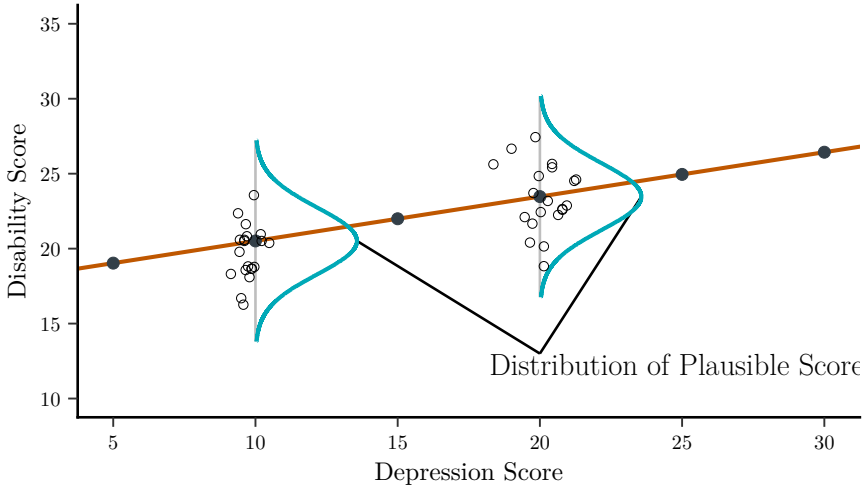
Predicted Scores on Outcome



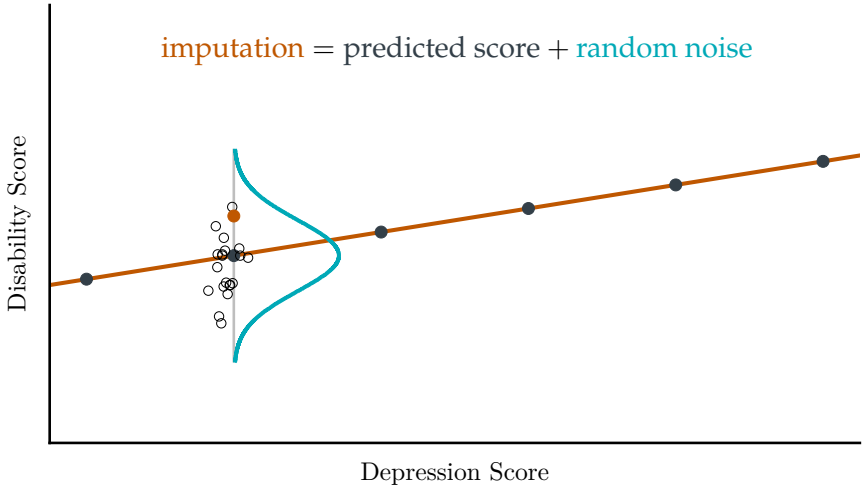
Distribution of Imputations



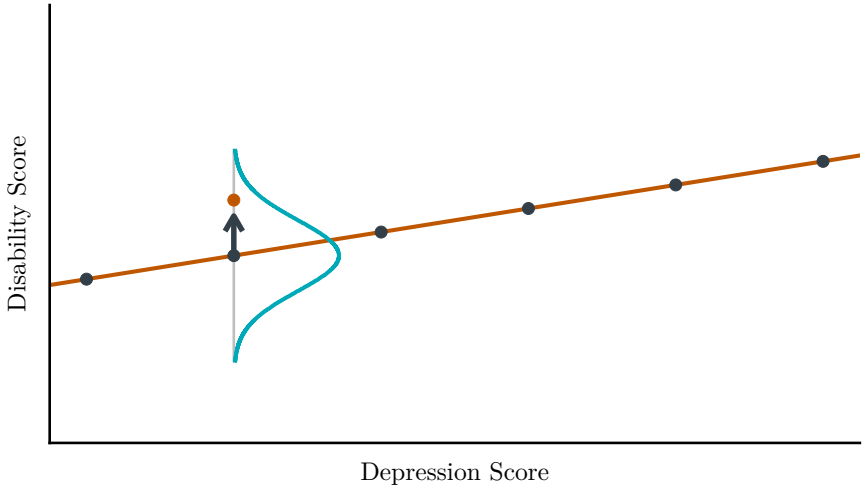
Distribution of Imputations



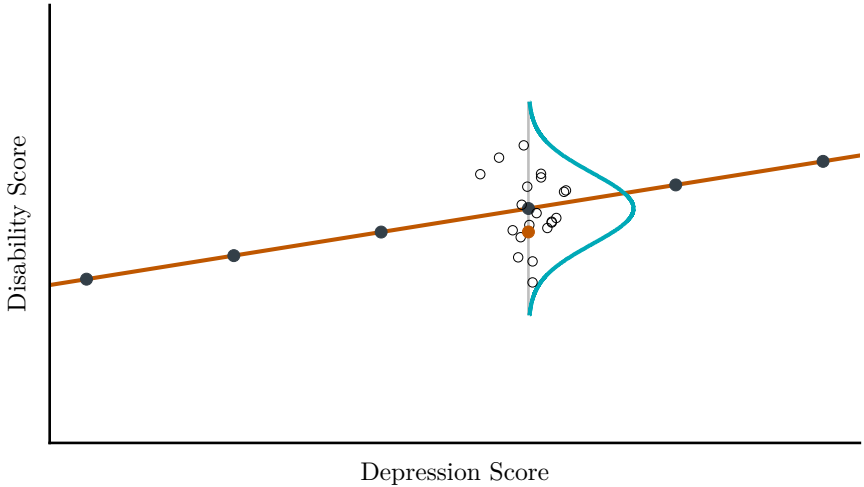
Sampling an Imputation



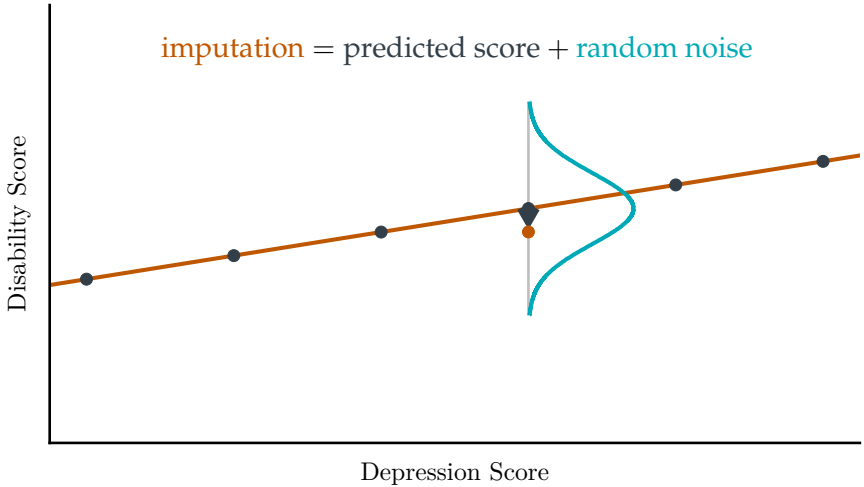
Sampling an Imputation



Sampling an Imputation



Sampling an Imputation



Incomplete Predictors

- Incomplete predictor variables require a model and distribution.
- We must model the multivariate distribution of our outcome and predictors.
- Using a pure multivariate distribution can ***misspecify*** the data distribution and cause biased estimates.

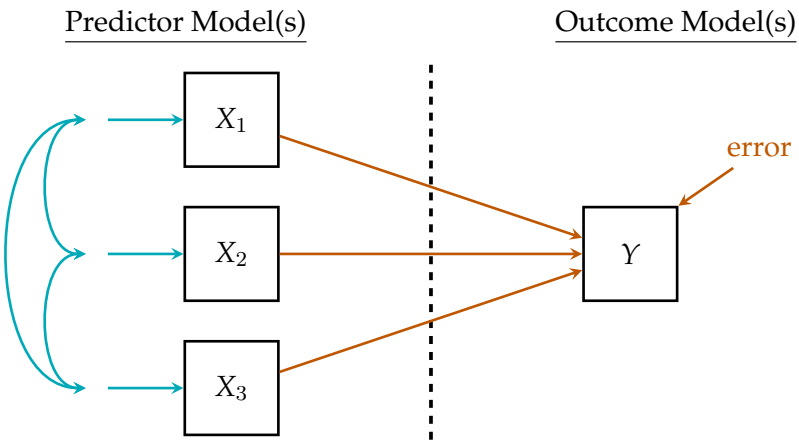
Factored Regression Modeling

Blimp's Modeling Framework

Factored Regression modeling invokes separate regression models (and distributions) for incomplete predictors and outcomes.

$$f\left(\text{outcomes}, \text{predictors}\right) = f\left(\text{outcomes} \mid \text{predictors}\right) \times f\left(\text{predictors}\right)$$

$$f(\text{outcomes}, \text{predictors}) = f(\text{outcomes} \mid \text{predictors}) \times f(\text{predictors})$$



Returning to Simple Regression Example...

Factorization:

$$f(\texttt{disab}, \texttt{depress}) = f\big(\texttt{disab} \mid \texttt{depress}\big) \times f(\texttt{depress})$$

Fitted Models:

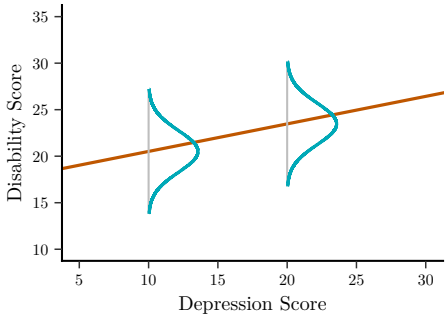
$$f\big(\texttt{disab} \mid \texttt{depress}\big) \rightarrow \qquad \texttt{disab}_i = \beta_0 + \beta_1 \left(\texttt{depress}_i\right) + e_i$$

$$f(\texttt{depress}) \rightarrow \qquad \texttt{depress}_i = \gamma_0 + r_i$$

Distributions...

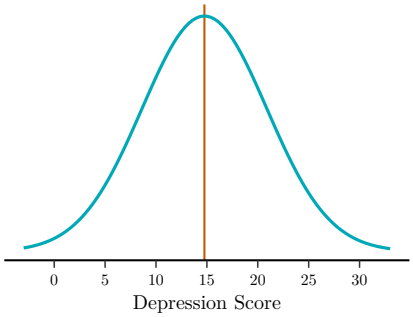
$$f(\texttt{disab} \mid \texttt{depress}) \rightarrow$$

$$\texttt{disab}_i = \beta_0 + \beta_1 (\texttt{depress}_i) + e_i$$



$$f(\texttt{depress}) \rightarrow$$

$$\texttt{depress}_i = \gamma_0 + r_i$$



Imputing Incomplete Predictors

- The missing predictor appears in two models (at least):
  - A predictor in the focal model
  - As an outcome in it's own model

- The distribution for the missing predictor is a composite of the two distributions:

$$f(\texttt{depress} \mid \texttt{disab}) \propto f(\texttt{disab} \mid \texttt{depress}) \times f(\texttt{depress})$$

- The distributions must be analytically solved based on the factorization.

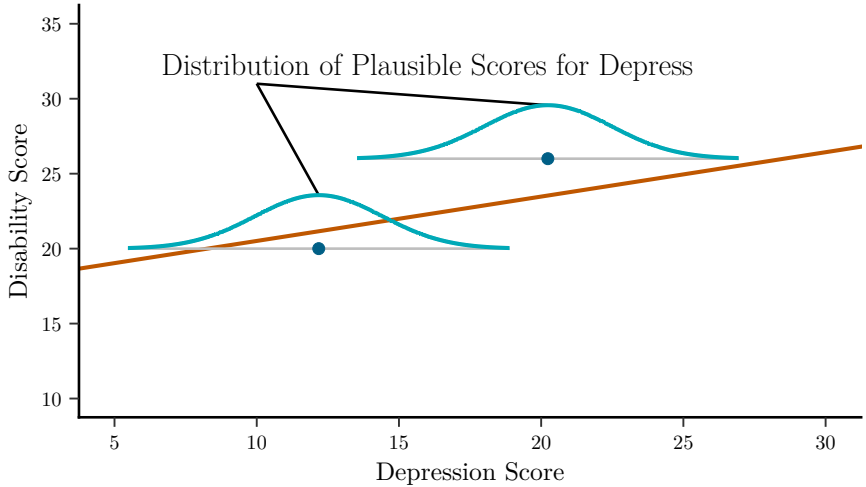
$$f\big(\texttt{depress} \mid \texttt{disab}\big) \propto f\big(\texttt{disab} \mid \texttt{depress}\big) \times f(\texttt{depress})$$

- The software will do this for you, but as an example:

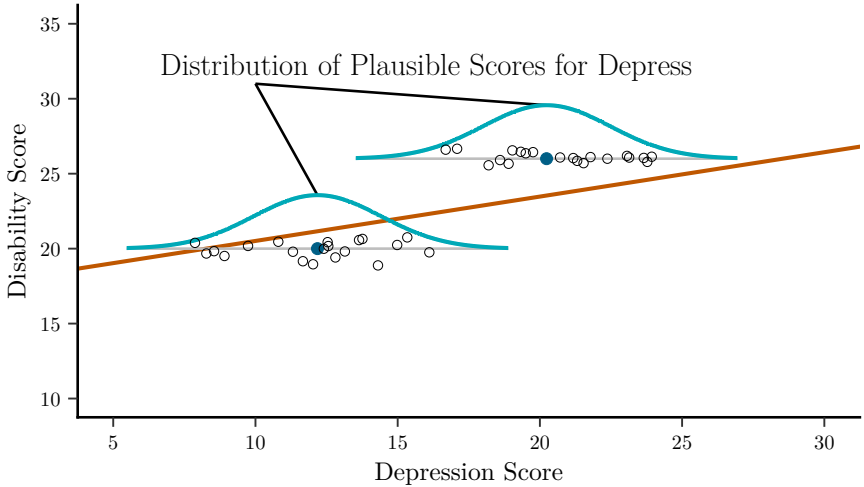
$$\hat{\texttt{depress}}_i = \frac{\sigma_r^2 \beta_1 \left(\texttt{disab}_i - \beta_0\right) + \sigma_e^2 \gamma_0}{\sigma_r^2 \beta_1^2 + \sigma_e^2}$$

$$\sigma^2_{\texttt{depress}|\texttt{disab}} = \frac{\sigma_e^2 \sigma_r^2}{\sigma_r^2 \beta_1^2 + \sigma_e^2}$$

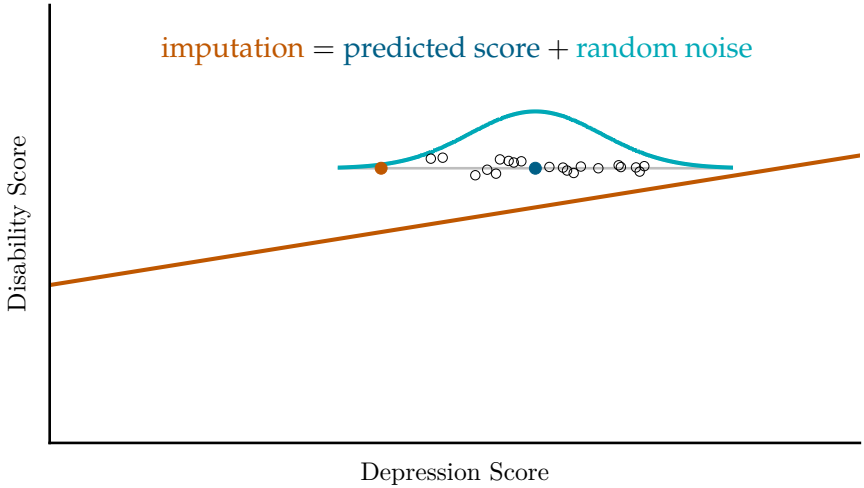
Distribution of Imputations for Predictor



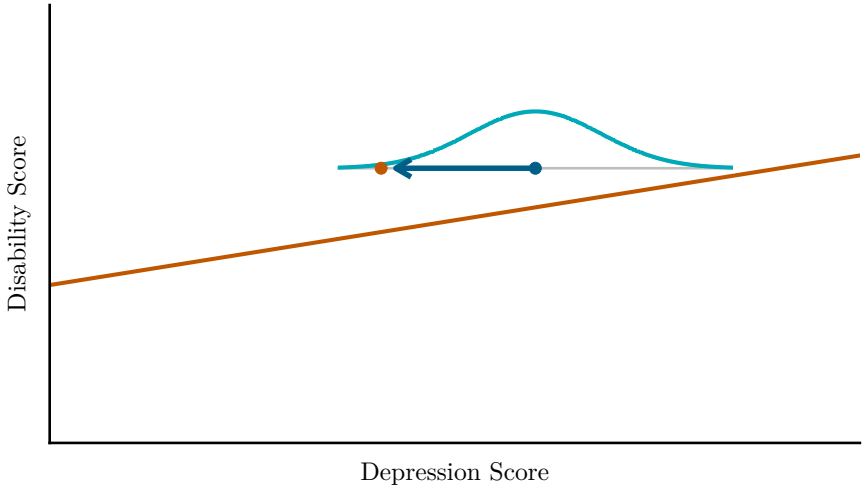
Distribution of Imputations



Sampling an Imputation for Predictor



Sampling an Imputation for Predictor



Fitting Regression Models in Blimp

Workshop Data

Chronic Pain Data – Variable Definitions

Name	Definition	Missing %	Range
`disab`	Psychosocial disability composite	9.1	10 to 34
`depress`	Depression composite score	13.5	7 to 28
`control`	Perceived pain control composite	0.0	6 to 30
`severity`	Severe pain dummy code	7.3	0 or 1
`male`	Biological sex dummy Code	0.0	0 or 1

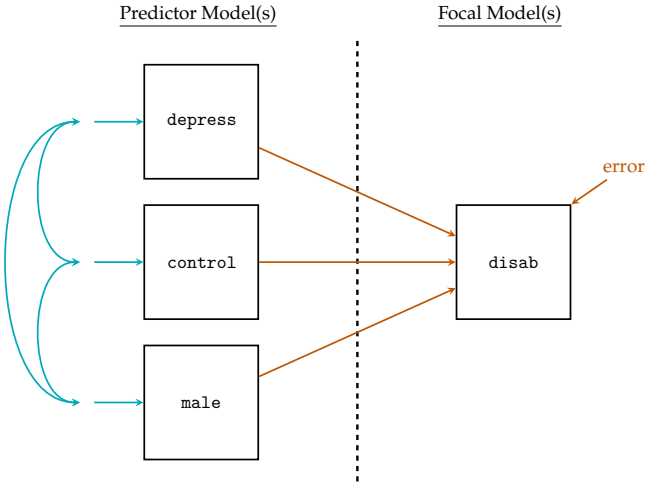
Analysis Model

- A study that is investigating the relationship of psychological disability predicted by depression scores, perceived control over pain, and biological sex.

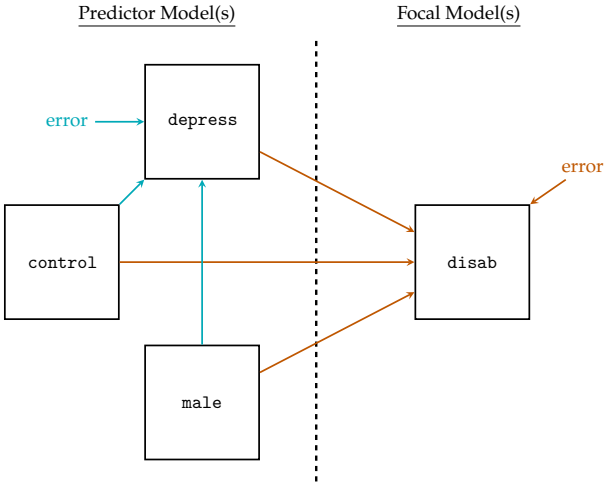
$$\texttt{disab}_i = \beta_0 + \beta_1 \left(\texttt{depress}_i\right) + \beta_2 \left(\texttt{control}_i\right) + \beta_3 \left(\texttt{male}_i\right) + e_i$$

`disab` and `depress` have missing values that must be handled.

$$f(\texttt{disab} \mid \texttt{depress}, \texttt{control}, \texttt{male}) \times f(\texttt{depress}, \texttt{control}, \texttt{male})$$



$$f(\texttt{disab} \mid \texttt{depress}, \texttt{control}, \texttt{male}) \times f(\texttt{depress} \mid \texttt{control}, \texttt{male})$$



Blimp Script for Multiple Regression

```
DATA: pain.dat;              # Read Data in
VARIABLES:                   # List Variable Names
    id txgrp male age edugroup workhrs
    exercise pain severity anxiety stress
    control depress interfere disab
    dep1:dep7 interf1:interf6 disab1:disab6;
MISSING: 999;                # Missing data code

# Specify Regression Model
MODEL:
    focal:     disab ~ depress control male;
    predictor: depress ~ control male;

# Specify the MCMC sampler parameters
SEED: 19723;                 # Set a prng seed
BURN: 2500;                  # Set number of burn iterations
ITERATIONS: 10000;           # Set number of analysis iterations
CHAINS: 4;                   # Specify number independent of chains
```

DATA and VARIABLES

```
DATA: pain.dat;                 # Read Data in
VARIABLES:                      # List Variable Names
    id txgrp male age edugroup workhrs
    exercise pain severity anxiety stress
    control depress interfere disab
    dep1:dep7 interf1:interf6 disab1:disab6;
MISSING: 999;                   # Missing data code
```

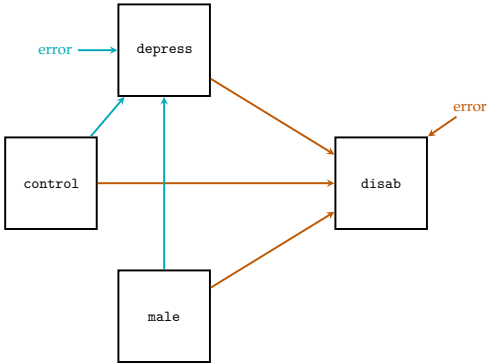
MODEL Specification

```
# Specify Regression Model
MODEL:
    focal:     disab ~ depress control male;
    predictor: depress ~ control male;
```

MODEL Specification

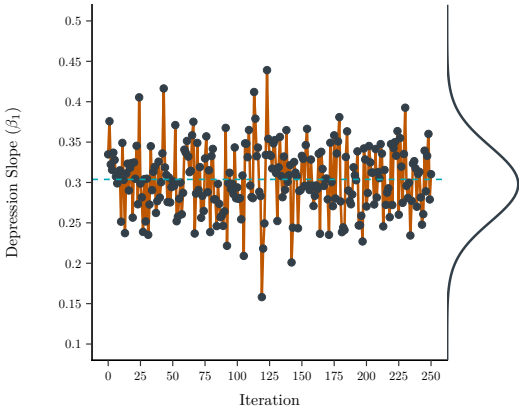
```
# Specify Regression Model
MODEL:
    focal:
    disab ~ depress control male;
    predictor:
    depress ~ control male;
```

Path Diagram



MCMC Estimation

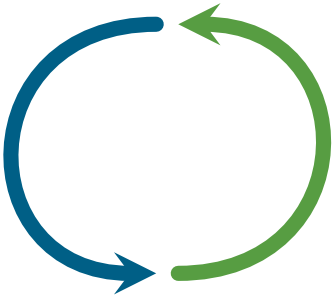
- MCMC uses a computer simulation to sample parameters from their posterior distribution.
- Each individual estimate itself is meaningless, but the entire sample of estimates can be used to characterize the distribution.



Understanding MCMC Specification

Within Each Iteration

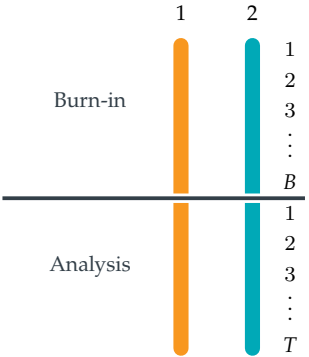
Estimate Regression Models



Fill-in Missing Values

Across the MCMC Run

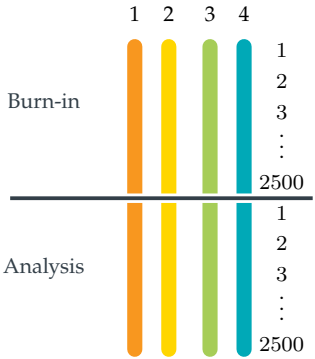
MCMC Chains



MCMC Sampler Specification

```
## Specify the MCMC sampler
# Set a prng seed
SEED: 19723;
# Set number of burn iterations
BURN: 2500;
# Set number of analysis iterations
ITERATIONS: 10000;
# Specify number of chains
CHAINS: 4;
```

MCMC Chains



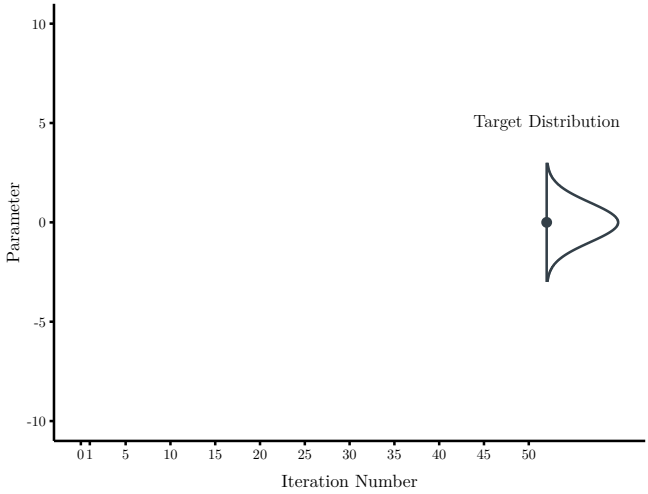
Understanding Blimp Output

MCMC Convergence

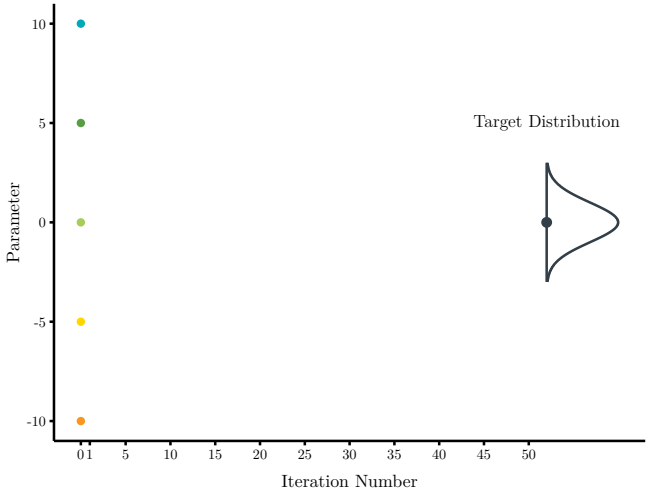
- MCMC "converges" when the posterior distributions are stationary.
- This means that parameter estimates oscillate around a stable mean, and the variation does not change with additional iterations.
- We should always set burn-in iterations greater than number of iterations needed for "convergence."

- The simulative algorithm MCMC uses requires monitoring of convergence.
- Convergence can loosely be thought of as determining if our algorithm has correctly drawn from the distribution it is trying to simulate from.
- We need to monitor convergence on every parameter in our model.

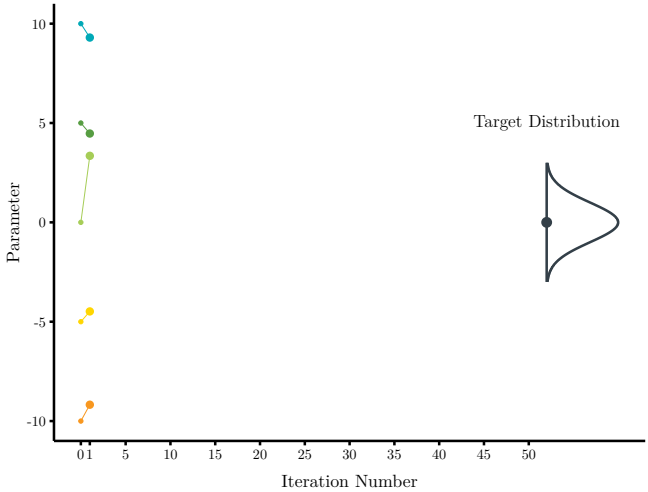
Distribution We are Trying To Simulate



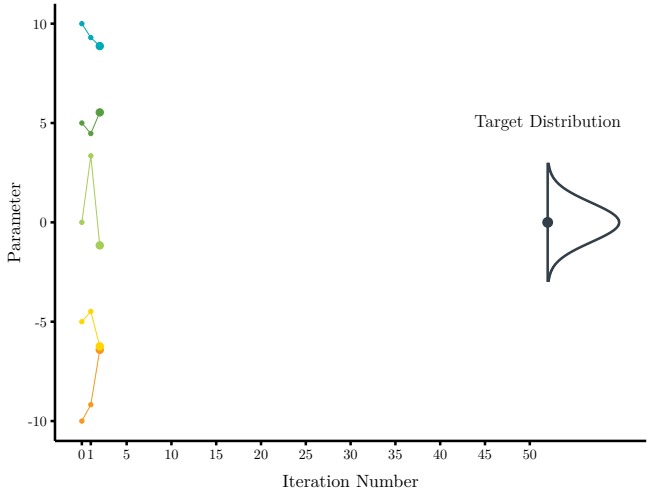
Arbitrary Starting Values for 5 Chains



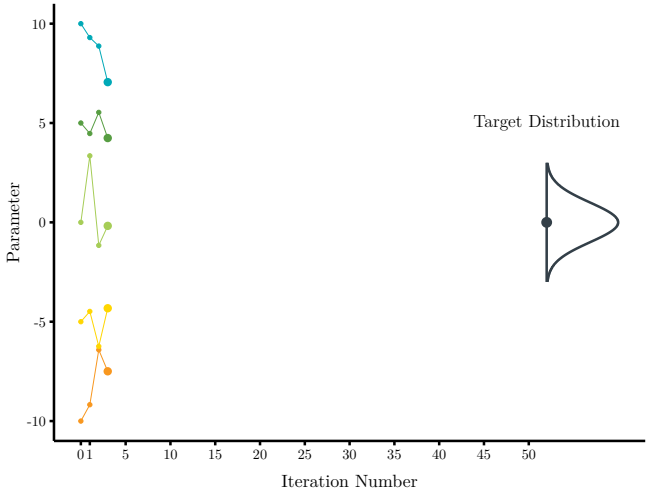
Iteration 1 for 5 Chains



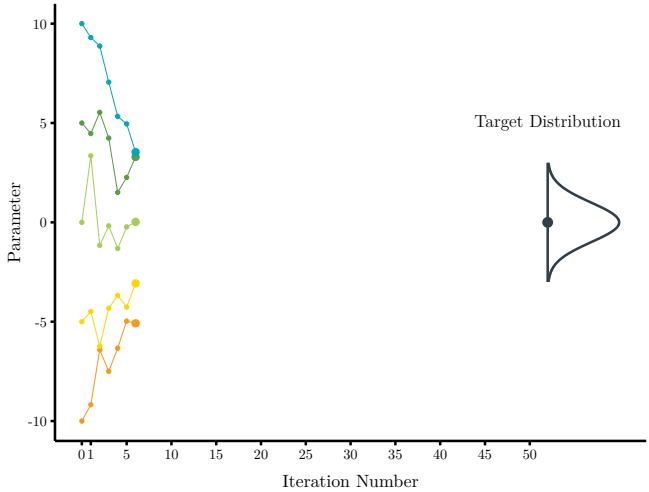
Iteration 2 for 5 Chains



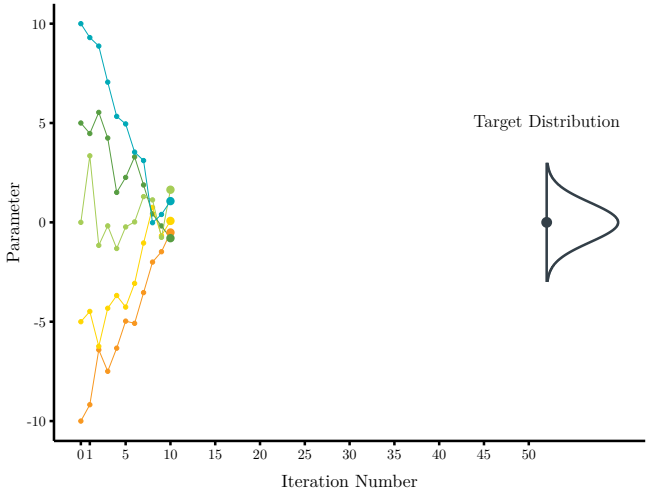
Iteration 3 for 5 Chains



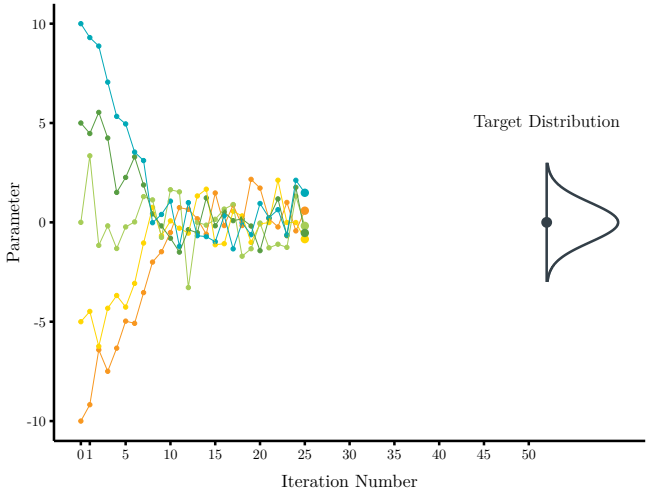
Iteration 6 for 5 Chains



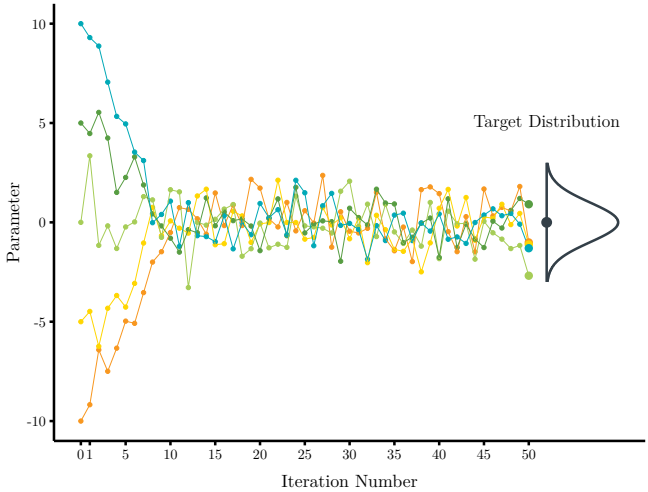
Iteration 10 for 5 Chains



Iteration 25 for 5 Chains



Iteration 50 for 5 Chains

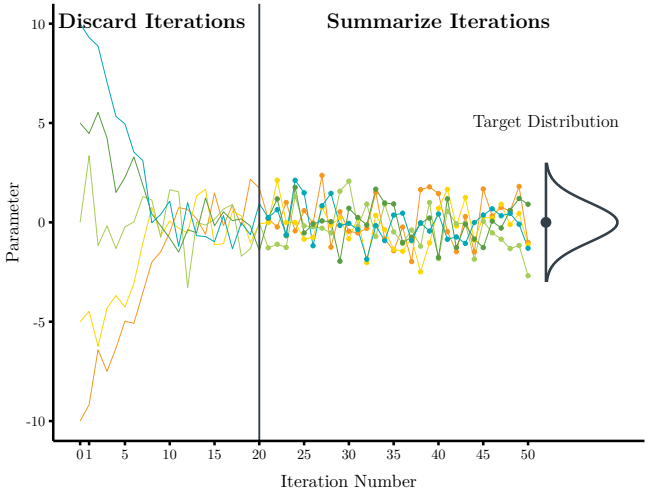


Bayesian Diagnostics

Diagnostics are used for two main purposes:

- To determine if we have issues estimating the model.
- To determine the appropriate burn-in iterations (i.e., how many iterations we will "throw away" before starting to sample from the distributions).

50 Iterations for 5 Chains

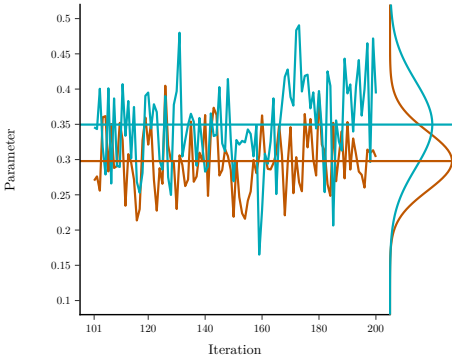


Potential Scale Reduction Factor

- The potential scale reduction factor (PSR, PSRF, Rhat, $\hat{R}$) compares parameter distributions generated from multiple unique MCMC processes.
- MCMC converges when the two chains give estimates with same mean and spread.
- A rule of thumb for PSRF is that convergence has been reached if the PSRF is constantly below 1.05.

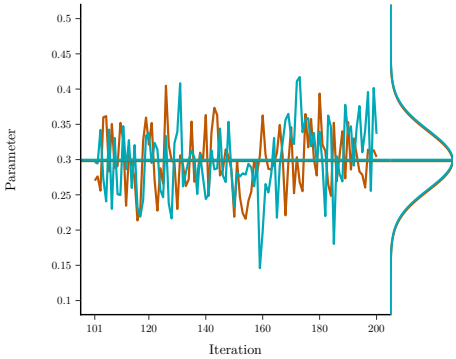
MCMC has not converged

($PSRF > 1.05$)



MCMC has converged

($PSRF < 1.05$)



Blimp PSRF Output

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 4 chains	Highest PSR	Parameter #
63 to 125	1.054	5
126 to 250	1.021	3
188 to 375	1.010	10
251 to 500	1.008	14
⋮		
1001 to 2000	1.002	4
1063 to 2125	1.002	16
1126 to 2250	1.002	12
1188 to 2375	1.002	12
1251 to 2500	1.002	12

Worst $PSRF < 1.05$

General Recommendations

- Look at the *PSRF* statistics. If the highest PSR falls below 1.05 or so and consistently stays below that then run a few extra iterations and use that as burn-in value.

- Most likely your model is not supported by your data (i.e., you do not have enough information to estimate the model). So simplify the model.

But...

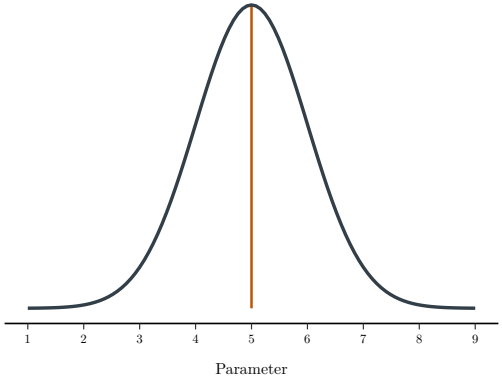
You can never run too many iterations!

Error on the side of caution always.

Posterior Mean and Median

- The posterior ***mean*** and ***median*** quantifies the parameter value most likely to be on average.

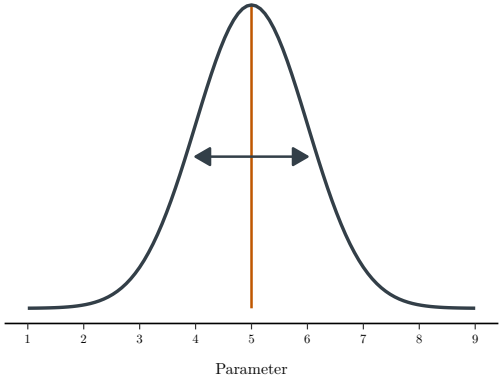
- They can be considered analog to point estimates in a standard Frequentist analysis.



Posterior Standard Deviation

- The posterior ***standard deviation*** quantifies the uncertainty we have in knowing the parameter value.

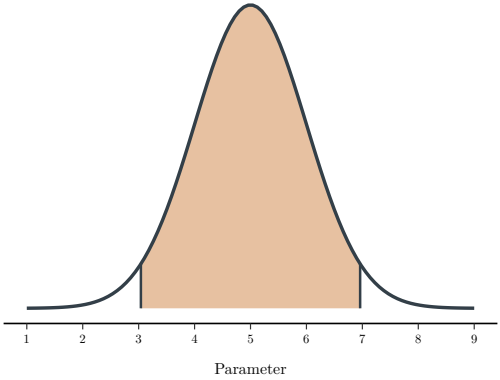
- This is similar to a Frequentist standard error but without reference to repeated sampling.



95% Credible Intervals

- The posterior ***95% credible interval*** quantifies the limits that span 95% of the parameter's range.

- This is similar to a Frequentist confidence interval but references a range of highly plausible values.



Blimp Outcome Model Output for Focal Model

```
OUTCOME MODEL ESTIMATES:

  Summaries based on 10000 iterations using 4 chains.

  focal block:

Outcome Variable:  disab

Parameters                      Median      StdDev        2.5%       97.5%         PSR       N_Eff
-------------------------------------------------------------------------------------------------

Variances:
  Residual Var.                 17.338       1.596      14.573      20.768       1.001   7139.3422

Coefficients:
  Intercept                     21.978       1.548      18.965      24.995       1.000   6922.092
  depress                        0.257       0.048       0.163       0.351       1.000   6400.560
  control                       -0.178       0.054      -0.283      -0.071       1.000   7121.600
  male                          -0.461       0.542      -1.506       0.599       1.001   8150.781

Standardized Coefficients:
  depress                        0.342       0.060       0.219       0.452       1.001   6285.317
  control                       -0.199       0.059      -0.311      -0.080       1.000   7173.699
  male                          -0.048       0.056      -0.155       0.063       1.001   8143.181

Proportion Variance Explained
  by Coefficients                0.207       0.045       0.121       0.298       1.000   6372.151
  by Residual Variation          0.793       0.045       0.702       0.879       1.000   6372.151

-------------------------------------------------------------------------------------------------
```

Posterior Median

OUTCOME MODEL ESTIMATES:

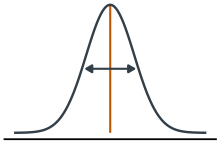
Summaries based on 10000 iterations using 4 chains.

focal block:

Outcome Variable: *disab*

Parameters	Median	StdDev	2.5%	97.5%	PSR	N_Eff
Variances:						
Residual Var.	17.338	1.596	14.573	20.768	1.001	7139.342
Coefficients:						
Intercept	21.978	1.548	18.965	24.995	1.000	6922.092
depress	0.257	0.048	0.163	0.351	1.000	6400.560
control	−0.178	0.054	−0.283	−0.071	1.000	7121.600
male	−0.461	0.542	−1.506	0.599	1.001	8150.781
Standardized Coefficients:						
depress	0.342	0.060	0.219	0.452	1.001	6285.317
control	−0.199	0.059	−0.311	−0.080	1.000	7173.699
male	−0.048	0.056	−0.155	0.063	1.001	8143.181
Proportion Variance Explained						
by Coefficients	0.207	0.045	0.121	0.298	1.000	6372.151
by Residual Variation	0.793	0.045	0.702	0.879	1.000	6372.151

Posterior Standard Deviation



OUTCOME MODEL ESTIMATES:

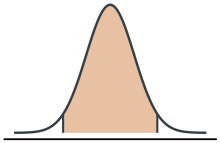
Summaries based on 10000 iterations using 4 chains.

focal block:

Outcome Variable: *disab*

Parameters	Median	StdDev	2.5%	97.5%	PSR	N_Eff
Variances:						
Residual Var.	17.338	1.596	14.573	20.768	1.001	7139.342
Coefficients:						
Intercept	21.978	1.548	18.965	24.995	1.000	6922.092
depress	0.257	0.048	0.163	0.351	1.000	6400.560
control	−0.178	0.054	−0.283	−0.071	1.000	7121.600
male	−0.461	0.542	−1.506	0.599	1.001	8150.781
Standardized Coefficients:						
depress	0.342	0.060	0.219	0.452	1.001	6285.317
control	−0.199	0.059	−0.311	−0.080	1.000	7173.699
male	−0.048	0.056	−0.155	0.063	1.001	8143.181
Proportion Variance Explained						
by Coefficients	0.207	0.045	0.121	0.298	1.000	6372.151
by Residual Variation	0.793	0.045	0.702	0.879	1.000	6372.151

95% Credible Intervals



OUTCOME MODEL ESTIMATES:

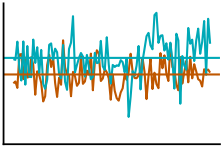
Summaries based on 10000 iterations using 4 chains.

focal block:

Outcome Variable: *disab*

Parameters	Median	StdDev	2.5%	97.5%	PSR	N_Eff
Variances:						
Residual Var.	17.338	1.596	14.573	20.768	1.001	7139.342
Coefficients:						
Intercept	21.978	1.548	18.965	24.995	1.000	6922.092
depress	0.257	0.048	0.163	0.351	1.000	6400.560
control	−0.178	0.054	−0.283	−0.071	1.000	7121.600
male	−0.461	0.542	−1.506	0.599	1.001	8150.781
Standardized Coefficients:						
depress	0.342	0.060	0.219	0.452	1.001	6285.317
control	−0.199	0.059	−0.311	−0.080	1.000	7173.699
male	−0.048	0.056	−0.155	0.063	1.001	8143.181
Proportion Variance Explained						
by Coefficients	0.207	0.045	0.121	0.298	1.000	6372.151
by Residual Variation	0.793	0.045	0.702	0.879	1.000	6372.151

Post-Burn Diagnostic Output



OUTCOME MODEL ESTIMATES:

Summaries based on 10000 iterations using 4 chains.

focal block:

Outcome Variable: *disab*

Parameters	Median	StdDev	2.5%	97.5%	PSR	N_Eff
Variances:						
Residual Var.	17.338	1.596	14.573	20.768	1.001	7139.342
Coefficients:						
Intercept	21.978	1.548	18.965	24.995	1.000	6922.092
depress	0.257	0.048	0.163	0.351	1.000	6400.560
control	−0.178	0.054	−0.283	−0.071	1.000	7121.600
male	−0.461	0.542	−1.506	0.599	1.001	8150.781
Standardized Coefficients:						
depress	0.342	0.060	0.219	0.452	1.001	6285.317
control	−0.199	0.059	−0.311	−0.080	1.000	7173.699
male	−0.048	0.056	−0.155	0.063	1.001	8143.181
Proportion Variance Explained						
by Coefficients	0.207	0.045	0.121	0.298	1.000	6372.151
by Residual Variation	0.793	0.045	0.702	0.879	1.000	6372.151

Variances and Covariances

OUTCOME MODEL ESTIMATES:

Summaries based on 10000 iterations using 4 chains.

focal block:

Outcome Variable: *disab*

Parameters	Median	StdDev	2.5%	97.5%	PSR	N_Eff
Variances:						
Residual Var.	17.338	1.596	14.573	20.768	1.001	7139.342
Coefficients:						
Intercept	21.978	1.548	18.965	24.995	1.000	6922.092
depress	0.257	0.048	0.163	0.351	1.000	6400.560
control	−0.178	0.054	−0.283	−0.071	1.000	7121.600
male	−0.461	0.542	−1.506	0.599	1.001	8150.781
Standardized Coefficients:						
depress	0.342	0.060	0.219	0.452	1.001	6285.317
control	−0.199	0.059	−0.311	−0.080	1.000	7173.699
male	−0.048	0.056	−0.155	0.063	1.001	8143.181
Proportion Variance Explained						
by Coefficients	0.207	0.045	0.121	0.298	1.000	6372.151
by Residual Variation	0.793	0.045	0.702	0.879	1.000	6372.151

Coefficients

OUTCOME MODEL ESTIMATES:

Summaries based on 10000 iterations using 4 chains.

focal block:

Outcome Variable: *disab*

Parameters	Median	StdDev	2.5%	97.5%	PSR	N_Eff
Variances:						
Residual Var.	17.338	1.596	14.573	20.768	1.001	7139.342
Coefficients:						
Intercept	21.978	1.548	18.965	24.995	1.000	6922.092
depress	0.257	0.048	0.163	0.351	1.000	6400.560
control	−0.178	0.054	−0.283	−0.071	1.000	7121.600
male	−0.461	0.542	−1.506	0.599	1.001	8150.781
Standardized Coefficients:						
depress	0.342	0.060	0.219	0.452	1.001	6285.317
control	−0.199	0.059	−0.311	−0.080	1.000	7173.699
male	−0.048	0.056	−0.155	0.063	1.001	8143.181
Proportion Variance Explained						
by Coefficients	0.207	0.045	0.121	0.298	1.000	6372.151
by Residual Variation	0.793	0.045	0.702	0.879	1.000	6372.151

Standardized Coefficients

OUTCOME MODEL ESTIMATES:

Summaries based on 10000 iterations using 4 chains.

focal block:

Outcome Variable: *disab*

Parameters	Median	StdDev	2.5%	97.5%	PSR	N_Eff
Variances:						
Residual Var.	17.338	1.596	14.573	20.768	1.001	7139.342
Coefficients:						
Intercept	21.978	1.548	18.965	24.995	1.000	6922.092
depress	0.257	0.048	0.163	0.351	1.000	6400.560
control	−0.178	0.054	−0.283	−0.071	1.000	7121.600
male	−0.461	0.542	−1.506	0.599	1.001	8150.781
Standardized Coefficients:						
depress	0.342	0.060	0.219	0.452	1.001	6285.317
control	−0.199	0.059	−0.311	−0.080	1.000	7173.699
male	−0.048	0.056	−0.155	0.063	1.001	8143.181
Proportion Variance Explained						
by Coefficients	0.207	0.045	0.121	0.298	1.000	6372.151
by Residual Variation	0.793	0.045	0.702	0.879	1.000	6372.151

Effect Size Estimates

OUTCOME MODEL ESTIMATES:

Summaries based on 10000 iterations using 4 chains.

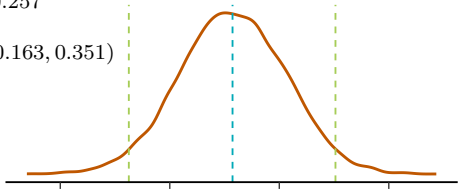
focal block:

Outcome Variable: *disab*

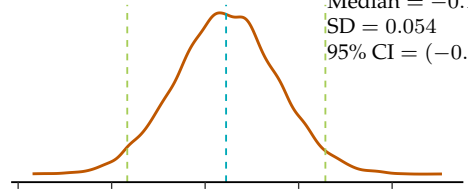
Parameters	Median	StdDev	2.5%	97.5%	PSR	N_Eff
Variances:						
Residual Var.	17.338	1.596	14.573	20.768	1.001	7139.342
Coefficients:						
Intercept	21.978	1.548	18.965	24.995	1.000	6922.092
depress	0.257	0.048	0.163	0.351	1.000	6400.560
control	−0.178	0.054	−0.283	−0.071	1.000	7121.600
male	−0.461	0.542	−1.506	0.599	1.001	8150.781
Standardized Coefficients:						
depress	0.342	0.060	0.219	0.452	1.001	6285.317
control	−0.199	0.059	−0.311	−0.080	1.000	7173.699
male	−0.048	0.056	−0.155	0.063	1.001	8143.181
Proportion Variance Explained						
by Coefficients	0.207	0.045	0.121	0.298	1.000	6372.151
by Residual Variation	0.793	0.045	0.702	0.879	1.000	6372.151

Posterior Distributions

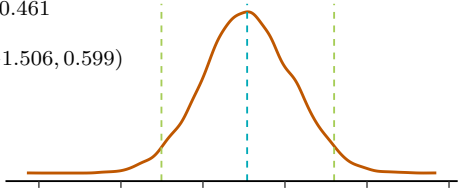
Median = 0.257
SD = 0.048
95% CI = (0.163, 0.351)



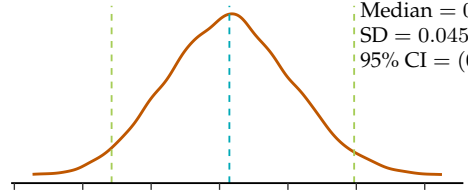
Median = −0.178
SD = 0.054
95% CI = (−0.283, −0.071)



Median = −0.461
SD = 0.542
95% CI = (−1.506, 0.599)



Median = 0.207
SD = 0.045
95% CI = (0.121, 0.298)



Analysis Model

- A study that is investigating the relationship of psychological disability predicted by depression scores, perceived control over pain, and biological sex.

$$\texttt{disab}_i = \beta_0 + \beta_1 \left(\texttt{depress}_i\right) + \beta_2 \left(\texttt{control}_i\right) + \beta_3 \left(\texttt{male}_i\right) + e_i$$

- ***Reminder!*** These are partial regression coefficients.

Example Interpretation

- For two participants who share the same biological sex and perceived control of pain, scoring one point higher on the depression scale is associated with a 0.257 increase on the psychological disability construct.
- The slope is "significantly different" from zero because zero is not within the 95% credible intervals (0.163, 0.351).

Do I have enough iterations?

- MCMC estimates are not independent across iterations.
- The effective sample size (`N_EFF`) estimates the number of independent samples that are used in the posterior summaries.
- The literature recommends at least 100 independent MCMC samples per parameter (Gelman et al., 2014; p. 267).

Effective Sample Size in Blimp

OUTCOME MODEL ESTIMATES:

Summaries based on 10000 iterations using 4 chains.

focal block:

Outcome Variable: *disab*

Parameters	Median	StdDev	2.5%	97.5%	PSR	N_Eff
Variances:						
Residual Var.	17.338	1.596	14.573	20.768	1.001	7139.342
Coefficients:						
Intercept	21.978	1.548	18.965	24.995	1.000	6922.092
depress	0.257	0.048	0.163	0.351	1.000	6400.560
control	−0.178	0.054	−0.283	−0.071	1.000	7121.600
male	−0.461	0.542	−1.506	0.599	1.001	8150.781
Standardized Coefficients:						
depress	0.342	0.060	0.219	0.452	1.001	6285.317
control	−0.199	0.059	−0.311	−0.080	1.000	7173.699
male	−0.048	0.056	−0.155	0.063	1.001	8143.181
Proportion Variance Explained						
by Coefficients	0.207	0.045	0.121	0.298	1.000	6372.151
by Residual Variation	0.793	0.045	0.702	0.879	1.000	6372.151

- *All parameters effective sample size are above 100.*

Estimator Comparison

- Two approaches provide numerically equivalent results and inferences.

	Bayesian				Frequentist (FIML)			
	Median	SD	CI 2.5%	CI 97.5%	Estimate	SE	CI 2.5%	CI 97.5%
Intercept (β_0)	21.98	1.55	18.97	25.00	21.86	1.58	18.77	24.96
Depression Slope (β_1)	0.26	0.05	0.16	0.35	0.26	0.05	0.16	0.35
Control Slope (β_2)	−0.18	0.05	−0.28	−0.07	−0.17	0.06	−0.28	−0.06
Male Slope (β_3)	−0.46	0.54	−1.51	0.60	−0.53	0.56	−1.64	0.57
Residual Var. (σ_e^2)	17.34	1.60	14.57	20.77	16.74	1.61	13.58	19.90

What about the predictor model?

- Blimp also supplies output for the specified predictor model.
- These parameter estimates can be considered nuisance parameters—that is, parameters we don't generally care about.
- We still need to monitor their convergence, including looking at PSRF and effective sample sizes.

Predictor Model Output

predictor block:

Outcome Variable: *depress*

Parameters	Median	StdDev	2.5%	97.5%	PSR	N_Eff
Variances:						
Residual Var.	34.596	3.247	28.955	41.666	1.001	7506.952
Coefficients:						
Intercept	22.470	1.574	19.369	25.526	1.000	7871.944
control	−0.390	0.072	−0.530	−0.249	1.000	7796.725
male	0.787	0.771	−0.695	2.317	1.001	7902.691
Standardized Coefficients:						
control	−0.327	0.055	−0.429	−0.212	1.001	7789.083
male	0.061	0.060	−0.055	0.178	1.001	7915.869
Proportion Variance Explained						
by Coefficients	0.116	0.037	0.053	0.195	1.001	7753.436
by Residual Variation	0.884	0.037	0.805	0.947	1.001	7753.436

Bayesian Wald Test

- Asparouhov and Muthén (2021) proposed a Bayesian Wald test that mimics familiar likelihood-based Wald tests.

$$T = (\boldsymbol{\theta} - \boldsymbol{\theta}_0)\, \boldsymbol{\Sigma}_{\theta}^{-1}\, (\boldsymbol{\theta} - \boldsymbol{\theta}_0)$$

- T is the sum of squared standardized differences (chi-square metric) between the posterior means and null hypothesis.

Adding TEST Command

```
# Specify Regression Model
MODEL:
    # Label coefficients with @
    focal:     disab ~ depress@b1 control@b2 male@b3;
    predictor: depress ~ control male;

# Perform Wald Tests: 4 total tests
TEST: b1 = 0; # Test of a single slope
TEST: b2 = 0; # test of a single slope
TEST: b3 = 0; # Test of a single slope
TEST: b1:b3 = 0; # Ominbus test of all 3 coefficients
```

Wald Test Output

```
MODEL FIT:

                                    ⋮

WALD TESTS (Asparouhov & Muthén, 2021)

Test #1

  Full:
    [1]  disab ~ Intercept depress@b1 control@b2 male@b3

  Restricted:
    [1]  disab ~ Intercept depress@b1 control@b2 male@b3

  Constraints in Restricted:
    [1]  b1 = 0


  Wald Statistic (Chi-Square)               28.271
  Number of Parameters Tested (df)               1
  Probability                                0.000
```

Wald Test #2 Output

```
Test #

  Full:
    [1]  disab ~ Intercept depress@b1 control@b2 male@b3

  Restricted:
    [1]  disab ~ Intercept depress@b1 control@b2 male@b3

  Constraints in Restricted:
    [1]  b2 = 0


  Wald Statistic (Chi-Square)               10.729
  Number of Parameters Tested (df)               1
  Probability                                0.001
```

Wald Test #3 Output

```
Test #3

  Full:
    [1]  disab ~ Intercept depress@b1 control@b2 male@b3

  Restricted:
    [1]  disab ~ Intercept depress@b1 control@b2 male@b3

  Constraints in Restricted:
    [1]  b3 = 0


  Wald Statistic (Chi-Square)                0.701
  Number of Parameters Tested (df)               1
  Probability                                0.402
```

Wald Test #4 Output

```
Test #4

  Full:
    [1]  disab ~ Intercept depress@b1 control@b2 male@b3

  Restricted:
    [1]  disab ~ Intercept depress@b1 control@b2 male@b3

  Constraints in Restricted:
    [1]  b1 = 0
    [2]  b2 = 0
    [3]  b3 = 0


  Wald Statistic (Chi-Square)               55.802
  Number of Parameters Tested (df)               3
  Probability                                0.000
```

Reporting Template – Diagnostics

We used Bayesian methods in Blimp 3.2 (Keller & Enders, 2021) to treat missing values under the assumption that missingness is random after conditioning on the observed data. Potential scale reduction factor convergence diagnostics (Gelman & Rubin, 1992) from a preliminary run indicated that a burn-in period of 2,500 iterations was sufficiently conservative. Based on this information, we used four MCMC chains with random starting values to generate posterior summaries consisting of 10,000 estimates following the initial burn-in period. We verified this setting was sufficient by examining the effective number of independent MCMC samples for each parameter, all of which were greater than the recommended value of 100 (Gelman et al., 2014, p. 287).

Reporting Template – Results

Table 1 displays the posterior summaries from the analysis using uninformative priors. The posterior medians and standard deviations are analogous to frequentist point estimates and standard errors, and the 95% credible interval limits are akin to confidence intervals. These quantities make no reference to repeated samples but instead convey parameter values that are consistent with the observed data. Given the same assumptions and data, Bayesian and likelihood-based missing data handling procedures are numerically equivalent (Enders, 2022).

Reporting Template – APA Table

Table 1

Parameter Summary from Bayesian Regression Analysis

	Median	SD	CI 2.5%	CI 97.5%
Intercept (β_0)	21.98	1.55	18.97	25.00
Depression Slope (β_1)	0.26	0.05	0.16	0.35
Control Slope (β_2)	−0.18	0.05	−0.28	−0.07
Male Slope (β_3)	−0.46	0.54	−1.51	0.60
R^2	0.21	0.05	0.12	0.30

Reporting Template – Results Continued

Collectively, the predictors explained approximately 21% of the variation in psychological disability scores. The Bayesian Wald test (Asparouhov & Muthén, 2021) of the full model was statistically significant, $\chi^2(3) = 55.80$, $p < 0.001$. While controlling for biological sex, depression scores exhibited a significant positive association with psychological disability scores ($\beta = 0.26$, 95% $CI = (0.16, 0.35)$), and the measure of perceived control of pain was inversely related to psychological disability ($\beta = -.18$, 95% $CI = -.28, -.07$)).

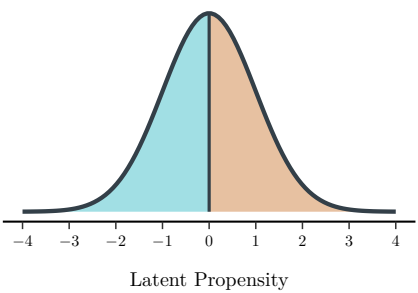
Incomplete Categorical Predictors in Blimp

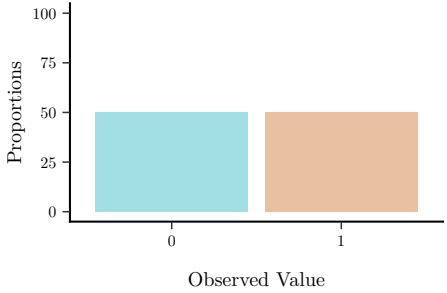
Factored Regression

- One advantage of Factored Regression is its ability to easily incorporate variables of different response metrics.
- By breaking up the complex multivariate distribution into easier chunks, it is easier to appropriately model the response metric.
- Blimp can incorporate binary, ordinal, and nominal predictor variables with minor changes to the Blimp syntax.

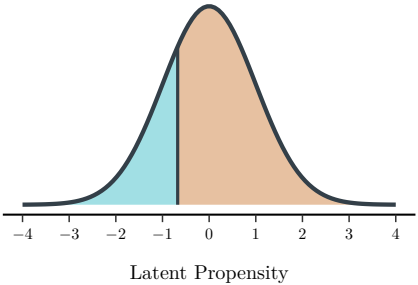
Latent Propensity Distribution

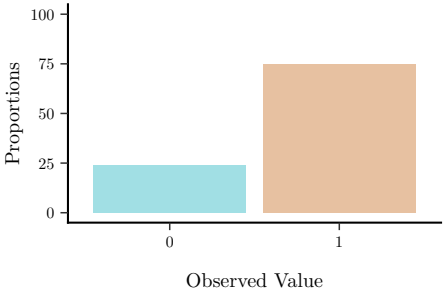
- A threshold parameter divides the latent distribution into segments, with areas under the curve matching the observed proportions.



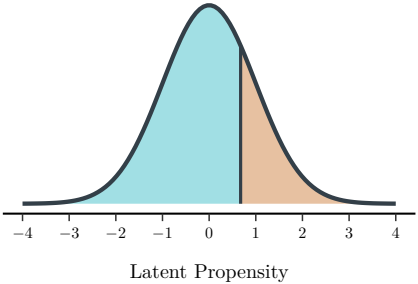


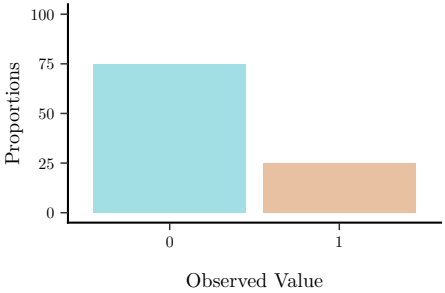
- When the threshold changes (or intercept) changes, then the proportion changes as well.





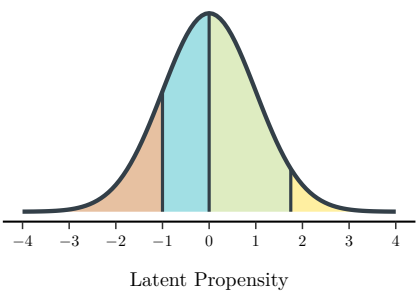
- When the threshold changes (or intercept) changes, then the proportion changes as well.

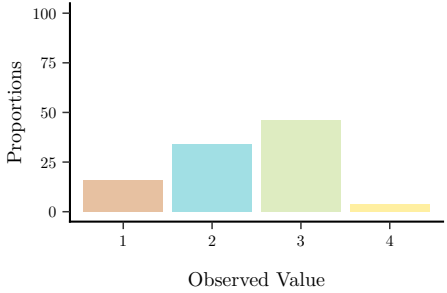




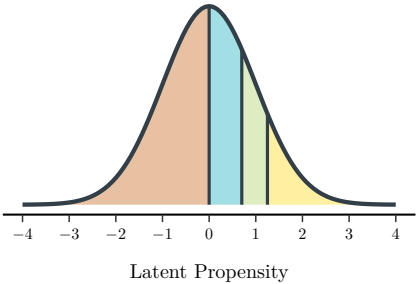
Latent Propensity Distribution

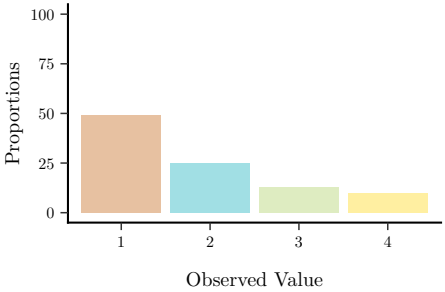
- Ordinal variables are defined by multiple thresholds to divide the propensity into multiple ordered categories.





- Changing the thresholds produce different distributions of the ordinal responses.



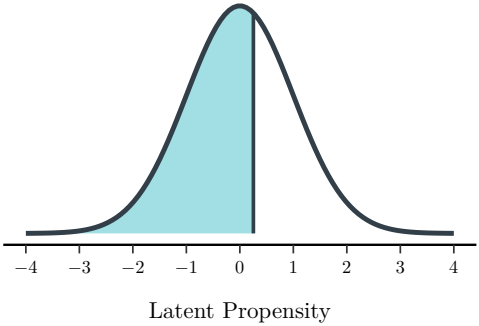


Treating the Latent Responses as Missing

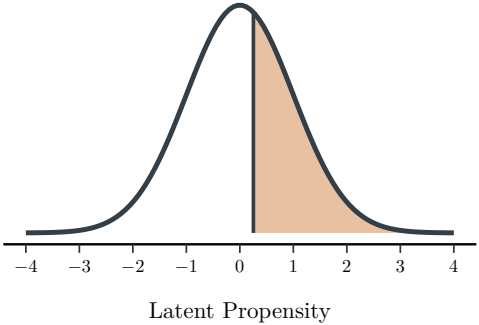
- The latent response scores can be considered missing data that must be imputed.

- MCMC treats the latent responses like any other incomplete variable with the caveat that we know the observed metric when the observed response is complete.

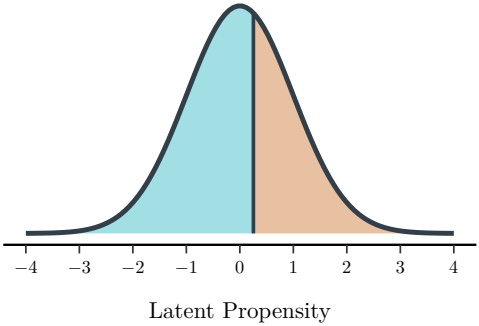
- For a binary variable, when the observed value is 0, we impute based on values below the threshold:



- When the observed value is 1, we impute based on values above the threshold:



- When the observed value is ***missing***, we impute based on values across the entire distribution:



Workshop Data

Chronic Pain Data – Variable Definitions

Name	Definition	Missing %	Range
`disab`	Psychosocial disability composite	9.1	10 to 34
`depress`	Depression composite score	13.5	7 to 28
`control`	Perceived pain control composite	0.0	6 to 30
`severity`	Severe pain dummy code	7.3	0 or 1
`male`	Biological sex dummy Code	0.0	0 or 1

Analysis Model

- A study that is investigating the relationship of psychological disability predicted by depression scores, coded severe pain (0 or 1), perceived control over pain, and biological sex.

$$\texttt{disab}_i = \beta_0 + \beta_1 \left(\texttt{depress}_i\right) + \beta_2 \left(\texttt{severity}_i\right) + \beta_3 \left(\texttt{control}_i\right) + \beta_4 \left(\texttt{male}_i\right) + e_i$$

Factored Regression Model

- With the additional incomplete variable (`severity`) our factorization of outcome and incomplete predictors is as follows:

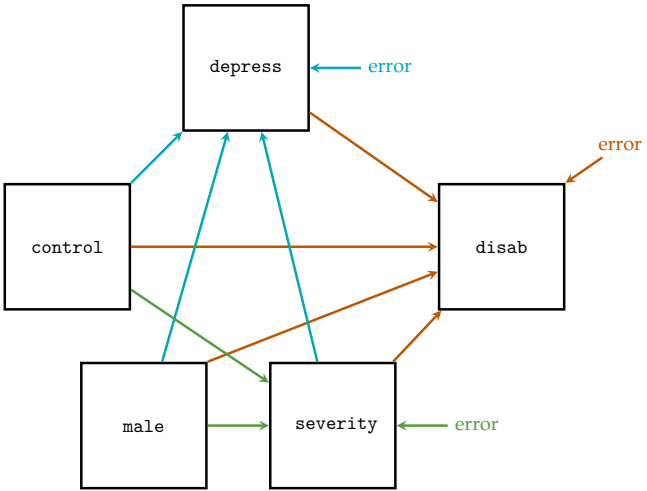
$$f(\texttt{disab} \mid \texttt{depress}, \texttt{severity}, \texttt{control}, \texttt{male}) \times$$
$$f(\texttt{depress}, \texttt{severity} \mid \texttt{control}, \texttt{male})$$

- We can further split the predictor model into two univariate models, giving us three full models we must specify:

$$
\begin{aligned}
&f\big(\texttt{disab} \mid \texttt{depress}, \texttt{severity}, \texttt{control}, \texttt{male}\big) \times \\
&f\big(\texttt{depress} \mid \texttt{severity}, \texttt{control}, \texttt{male}\big) \times \\
&f\big(\texttt{severity} \mid \texttt{control}, \texttt{male}\big)
\end{aligned}
$$

$$f(\texttt{disab} \mid \texttt{depress}, \texttt{control}, \texttt{male}) \times$$

$$f(\texttt{depress} \mid \texttt{severity}, \texttt{control}, \texttt{male}) \times f(\texttt{severity} \mid \texttt{control}, \texttt{male})$$



Blimp Script with Incomplete Binary Predictor

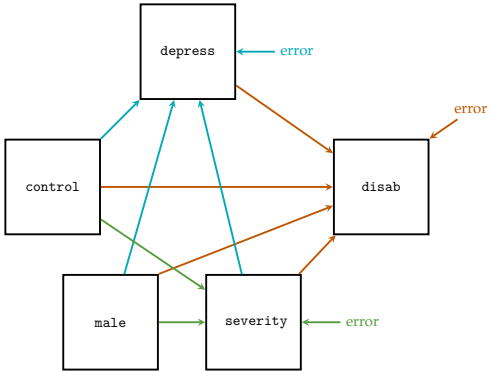
```
DATA: pain.dat;              # Read Data in
VARIABLES:                   # List Variable Names
    id txgrp male age edugroup workhrs exercise pain severity anxiety stress
    control depress interfere disab dep1:dep7 interf1:interf6 disab1:disab6;
MISSING: 999;                # Missing data code
ORDINAL: severity;           # Specify Binary Variables

# Specify Regression Model
MODEL:
    focal:
        disab ~ depress severity control male;

    predictor:
        depress ~ severity control male;
        severity ~ control male;

# Specify the MCMC sampler
```

⋮

MODEL Specification

```
# Specify Binary Variables
ORDINAL: severity;

# Specify Regression Model
MODEL:
    focal:
        disab ~ depress severity control male;

    predictor:
        depress ~ severity control male;
        severity ~ control male;
```

Path Diagram



Blimp PSRF Output

```
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 4 chains     Highest PSR   Parameter #
                          126 to 250             1.038            30
                          251 to 500             1.022            24
                          376 to 750             1.019            26
                          501 to 1000            1.012            26

                                  ⋮

                         2001 to 4000            1.003            25
                         2126 to 4250            1.003            25
                         2251 to 4500            1.003            25
                         2376 to 4750            1.003            25
Worst PSRF < 1.05        2501 to 5000            1.002            25
```

Blimp Outcome Model Output for Focal Model

focal block:

Outcome Variable: *disab*

Parameters	Median	StdDev	2.5%	97.5%	PSR	N_Eff
Variances:						
Residual Var.	16.927	1.597	14.236	20.525	1.000	7171.144
Coefficients:						
Intercept	21.487	1.541	18.436	24.561	1.001	7717.533
depress	0.238	0.048	0.142	0.331	1.001	6743.617
severity	1.571	0.611	0.368	2.764	1.001	6821.550
control	−0.158	0.054	−0.266	−0.050	1.001	7882.332
male	−0.789	0.555	−1.882	0.291	1.000	8026.602
Standardized Coefficients:						
depress	0.314	0.060	0.191	0.425	1.001	6461.162
severity	0.156	0.059	0.037	0.271	1.001	6793.303
control	−0.177	0.059	−0.290	−0.056	1.001	7970.693
male	−0.082	0.057	−0.194	0.030	1.000	8069.048
Proportion Variance Explained						
by Coefficients	0.230	0.045	0.145	0.321	1.000	6615.711
by Residual Variation	0.770	0.045	0.679	0.855	1.000	6615.711

Example Interpretation

- Comparing two participants who share the same on all other predictors (biological sex, perceived control of pain, and depression score), endorsing severe pain or not is associated with a 1.57 increase on the psychological disability construct.
- The slope is "significantly different" from zero because zero is not within the 95% credible intervals (0.368, 2.764).

Interaction Effects in Blimp

Moderated Regression (Interaction Effects)

- Moderation occurs when the magnitude of an association of a focal predictor (X) depends on a third variable (moderator; M).

- Moderated regressions answer the question, for whom does an effect apply?

Moderated Regression Model

A moderated regression adds the product of X and M as a predictor:

$$y_i = \beta_0 + \beta_1(x_i) + \beta_2(m_i) + \beta_3(x_i \times m_i) + e_i$$

β_3: Captures the change in the β_1 slope for every one-unit increase of M.

Incomplete Variables in Product Terms

- Factored regression readily handles interactive and other nonlinear effects
- Product terms appear as a deterministic function in the focal regression model.
- The models for the predictor variables remain unchanged.

Workshop Data

Chronic Pain Data – Variable Definitions

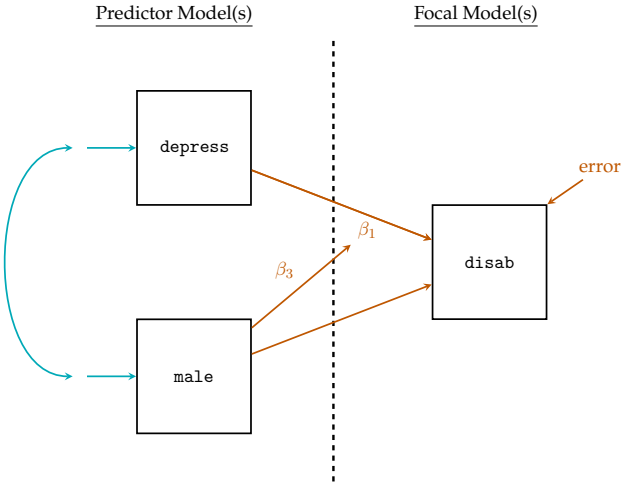
Name	Definition	Missing %	Range
`disab`	Psychosocial disability composite	9.1	10 to 34
`depress`	Depression composite score	13.5	7 to 28
`control`	Perceived pain control composite	0.0	6 to 30
`severity`	Severe pain dummy code	7.3	0 or 1
`male`	Biological sex dummy Code	0.0	0 or 1

Analysis Model

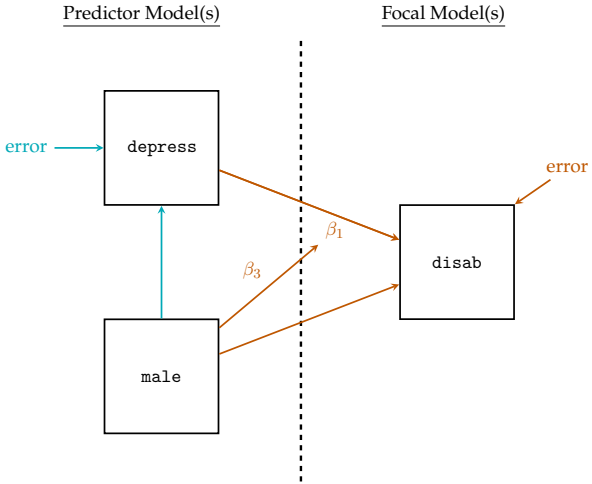
- A study that is investigating the relationship between psychological disability and depression scores moderated by biological sex.

$$\texttt{disab}_i = \beta_0 + \beta_1 \left(\texttt{depress}_i\right) + \beta_2 \left(\texttt{male}_i\right) + \beta_3 \left(\texttt{depress}_i \times \texttt{male}_i\right) + e_i$$

$$f(\texttt{disab} \mid \texttt{depress}, \texttt{male}) \times f(\texttt{depress}, \texttt{male})$$



$$f(\texttt{disab} \mid \texttt{depress}, \texttt{male}) \times f(\texttt{depress} \mid \texttt{male})$$



Factorization:

$$f\big(\texttt{disab} \mid \texttt{depress}, \texttt{male}\big) \times f\big(\texttt{depress} \mid \texttt{male}\big)$$

Fitted Models:

$$f\big(\texttt{disab} \mid \texttt{depress}, \texttt{male}\big) \rightarrow \qquad \texttt{disab}_i = \beta_0 + \beta_1\left(\texttt{depress}_i\right) + \beta_2\left(\texttt{male}_i\right) + \beta_3\left(\texttt{depress}_i \times \texttt{male}_i\right) + e_i$$

$$f\big(\texttt{depress} \mid \texttt{male}\big) \rightarrow \qquad \texttt{depress}_i = \gamma_0 + \gamma_1\left(\texttt{male}_i\right) + r_i$$

Blimp Script for Moderated Regression

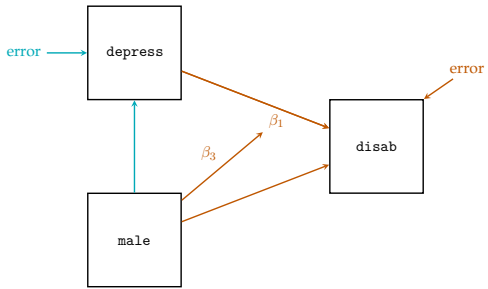
```
DATA: pain.dat;              # Read Data in
VARIABLES:                   # List Variable Names
    id txgrp male age edugroup workhrs exercise pain severity anxiety stress
    control depress interfere disab dep1:dep7 interf1:interf6 disab1:disab6;
ORDINAL: male;               # Specify Binary Variables
MISSING: 999;                # Missing data code

FIXED:  male;                # Specify variables with no missing
CENTER: depress;             # Center variables
# Specify Regression Model
MODEL:
    focal: disab ~ depress male depress*male;
    predictor: depress ~ male;

# Conditional Effects Analysis
SIMPLE: depress | male;

# Specify the MCMC sampler
```

:

MODEL Specification

```
# Specify variables with no missing
FIXED:  male;
# Center variables
CENTER: depress;
# Specify Regression Model
MODEL:
    focal:
        disab ~ depress male depress*male;
    predictor:
        depress ~ male;
```

Path Diagram



Blimp PSRF Output

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 4 chains	Highest PSR	Parameter #
126 to 250	1.041	12
251 to 500	1.015	17
376 to 750	1.013	17
501 to 1000	1.008	17
⋮		
2001 to 4000	1.002	2
2126 to 4250	1.002	3
2251 to 4500	1.002	17
2376 to 4750	1.002	17
2501 to 5000	1.002	17

Worst $PSRF < 1.05$

Blimp Outcome Model Output for Focal Model

focal block:

Outcome Variable: *disab*

Grand Mean Centered: depress

Parameters	Median	StdDev	2.5%	97.5%	PSR	N_Eff
Variances:						
Residual Var.	17.713	1.655	14.834	21.324	1.001	6395.983
Coefficients:						
Intercept	22.077	0.389	21.314	22.838	1.001	4422.551
depress	0.405	0.063	0.283	0.529	1.001	5230.217
male	−0.350	0.557	−1.434	0.731	1.000	7061.496
depress*male	−0.224	0.093	−0.406	−0.043	1.001	5492.627
Standardized Coefficients:						
depress	0.537	0.074	0.384	0.670	1.002	4953.854
male	−0.037	0.058	−0.148	0.076	1.000	7097.749
depress*male	−0.200	0.081	−0.356	−0.038	1.001	5449.779
Proportion Variance Explained						
by Coefficients	0.189	0.045	0.105	0.279	1.002	4701.979
by Residual Variation	0.811	0.045	0.721	0.895	1.002	4701.979

Conditional Effects Analysis (Simple Effects)

We can rearrange terms to give the conditional effect of a focal predictor (X) on a function of the moderator (M).

$$
\begin{aligned}
y_i &= \beta_0 + \beta_1 x_i + \beta_2 m_i + \beta_3 x_i m_i + e_i \\
&= (\beta_0 + \beta_2 m_i) + (\beta_1 x_i + \beta_3 x_i m_i) + e_i \\
&= (\beta_0 + \beta_2 m_i) + (\beta_1 + \beta_3 m_i)\, x_i + e_i \\
y_i &= \big(\text{simple intercept}\big) + \big(\text{simple slope}\big)\, x_i + e_i
\end{aligned}
$$

Conditional Effects Analysis in Blimp

- Compute for when $\texttt{male} = 0$:

$$\begin{aligned}\texttt{disab}_i &= (\beta_0 + \beta_2 \times 0) + (\beta_1 + \beta_3 \times 0)\,(\texttt{depress}_i - \mu_{\texttt{dep}}) + e_i \\ &= (\beta_0) + (\beta_1)\,(\texttt{depress} - \mu_{\texttt{dep}}) + e_i\end{aligned}$$

- Compute for when $\texttt{male} = 1$:

$$\begin{aligned}\texttt{disab}_i &= (\beta_0 + \beta_2 \times 1) + (\beta_1 + \beta_3 \times 1)\,(\texttt{depress}_i - \mu_{\texttt{dep}}) + e_i \\ &= (\beta_0 + \beta_2) + (\beta_1 + \beta_3)\,(\texttt{depress}_i - \mu_{\texttt{dep}}) + e_i\end{aligned}$$

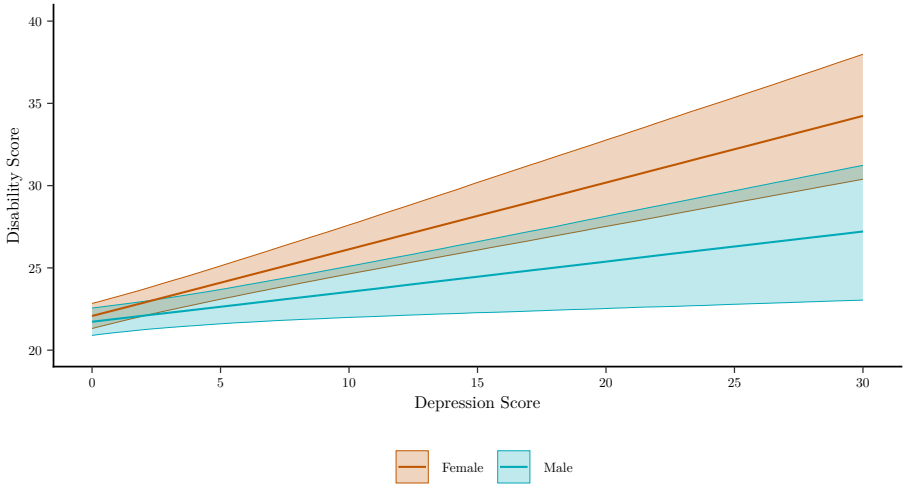
Simple Command

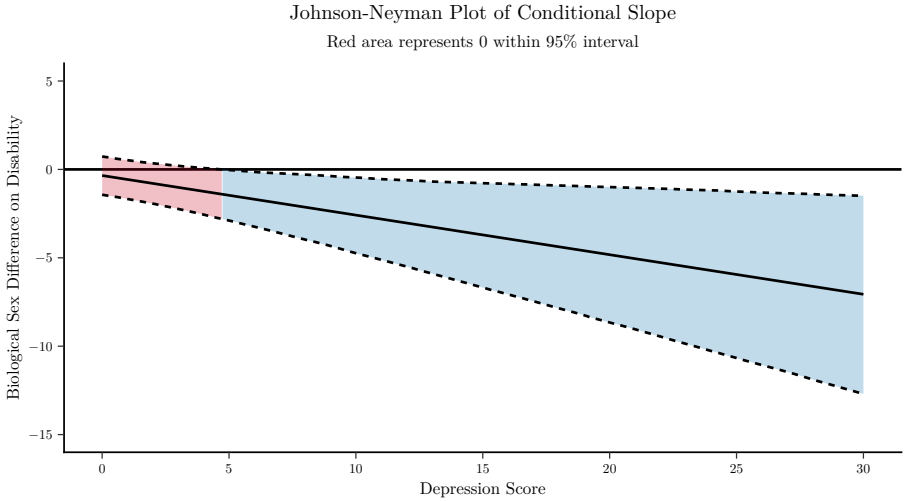
```
# Specify Regression Model
# Conditional Effects Analysis
SIMPLE: depress | male;

# Focal predictor given values of male values: 0 and 1
```

Blimp Output for SIMPLE Command

```
Conditional Effects                    Median      StdDev       2.5%       97.5%        PSR       N_Eff
-------------------------------------------------------------------------------------------------------
  depress | male @ 0
    Intercept                          22.077       0.389     21.314      22.838      1.001    4422.551
    Slope                               0.405       0.063      0.283       0.529      1.001    5230.217

  depress | male @ 1
    Intercept                          21.727       0.426     20.895      22.559      1.001    7120.648
    Slope                               0.182       0.070      0.045       0.316      1.001    5066.668

-------------------------------------------------------------------------------------------------------

NOTE: Intercepts are computed by setting all predictors
      not involved in the conditional effect to zero.
```

Plot of Conditional Effects




Johnson-Neyman Plot of Conditional Slope
Red area represents 0 within 95% interval
Biological Sex Difference on Disability
5
0
−5
−10
−15
0
5
10
15
20
25
30
Depression Score

Reporting Template – Diagnostics

We used Bayesian methods in Blimp 3.2 (Keller & Enders, 2021) to treat missing values under the assumption that missingness is random after conditioning on the observed data. Potential scale reduction factor convergence diagnostics (Gelman & Rubin, 1992) from a preliminary run indicated that a burn-in period of 5,000 iterations was sufficiently conservative. Based on this information, we used four MCMC chains with random starting values to generate posterior summaries consisting of 10,000 estimates following the initial burn-in period. We verified this setting was sufficient by examining the effective number of independent MCMC samples for each parameter, all of which were greater than the recommended value of 100 (Gelman et al., 2014, p. 287).

Reporting Template – Results

Table 1 displays the posterior summaries from the analysis using uninformative priors. The posterior medians and standard deviations are analogous to frequentist point estimates and standard errors, and the 95% credible interval limits are akin to confidence intervals. These quantities make no reference to repeated samples but instead convey parameter values that are consistent with the observed data. Given the same assumptions and data, Bayesian and likelihood-based missing data handling procedures are numerically equivalent (Enders, 2022).

Reporting Template – APA Table

Table 1

Parameter Summary from Bayesian Regression Analysis

	Median	SD	CI 2.5%	CI 97.5%
Intercept (β_0)	22.08	0.39	21.31	22.84
Depression Slope (β_1)	0.41	0.06	0.28	0.53
Male Slope (β_2)	-0.35	0.56	-1.434	0.731
Depression $\times$ Male (β_3)	-0.22	0.09	-0.41	0.04
R^2	0.19	0.05	0.11	0.28

Reporting Template – Results Continued

Collectively, the predictors explained approximately 19% of the variation in psychological disability scores. The 95% credible intervals for the interaction effect did not contain zero within the interval $(-0.41, -0.04)$, suggesting that biological sex moderates the association between psychological disability and depression composite scores. The negative interaction coefficient indicates that the strength of the positive association between psychological disability and depression decreases for males compared to females. Conducting a conditional effects analysis reveals that the strength of association is about half for men compared to women. Figure 1 displays the simple slopes for men and women.

Composite Scores with Item Level Missingness

- Often, psychological constructs are composites of multiple items summed or averaged together.

- When one item is missing, the entire composite score is considered missing. This can lead to a loss of information and reduced statistical power.

Workshop Data

Chronic Pain Data – Variable Definitions

Name	Definition	Missing %	Range
`disab`	Psychosocial disability composite	9.1	10 to 34
`depress`	Depression composite score	13.5	7 to 28
`male`	Biological sex dummy Code	0.0	0 or 1

Workshop Data – Depression Items

Chronic Pain Data – Variable Definitions

Name	Definition	Missing %	Range
`disab`	Psychosocial disability composite	9.1	10 to 34
`male`	Biological sex dummy Code	0.0	0 or 1
`dep1`	Couldn't experience any positive feelings at all	4.7	1 to 4
`dep2`	Difficult to work up the initiative to do things	2.2	1 to 4
`dep3`	I felt that I had nothing to look forward to	1.8	1 to 4
`dep4`	I felt down-hearted and blue	1.5	1 to 4
`dep5`	Unable to become enthusiastic about anything	2.2	1 to 4
`dep6`	I felt I wasn't worth much as a person	4.0	1 to 4
`dep7`	I felt that life was meaningless	2.9	1 to 4

Analysis Model

- A study that is investigating the relationship between psychological disability and depression composite moderated by biological sex.

$$\texttt{disab}_i = \beta_0 + \beta_1 \left(\texttt{dep}_i\right) + \beta_2 \left(\texttt{male}_i\right) + \beta_3 \left(\texttt{depress}_i \times \texttt{male}_i\right) + e_i$$

where...

$$\texttt{dep}_i = \texttt{dep1}_i + \texttt{dep2}_i + \texttt{dep3}_i + \ldots + \texttt{dep7}_i$$

Factored Regression Specification

- With the multiple items (`dep1` to `dep7`) our factorization of outcome and incomplete predictors is as follows:

$$f\big(\texttt{disab} \mid \texttt{dep1}, \texttt{dep2}, \texttt{dep3}, \texttt{dep4}, \ldots, \texttt{dep7}, \texttt{male}\big) \times$$
$$f\big(\texttt{dep1}, \texttt{dep2}, \texttt{dep3}, \texttt{dep4}, \ldots, \texttt{dep7} \mid \texttt{male}\big)$$

- We can further split the predictor model into two univariate models, giving us three full models we must specify:

$$
\begin{aligned}
f\big(\texttt{disab} \mid \texttt{dep1}, \texttt{dep2}, \texttt{dep3}, \texttt{dep4}, \ldots, \texttt{dep7}, \texttt{male}\big) \times \\
f\big(\texttt{dep1} \mid \texttt{dep2}, \texttt{dep3}, \texttt{dep4}, \ldots, \texttt{dep7}, \texttt{male}\big) \times \\
f\big(\texttt{dep2} \mid \texttt{dep3}, \texttt{dep4}, \ldots, \texttt{dep7}, \texttt{male}\big) \times \\
f\big(\texttt{dep3} \mid \texttt{dep4}, \ldots, \texttt{dep7}, \texttt{male}\big) \times \\
\vdots \\
f\big(\texttt{dep7} \mid \texttt{male}\big) \times
\end{aligned}
$$

Blimp Script for Moderated Regression with Sum Score

(Model Command Only)

```
FIXED:  male;           # Specify variables with no missing
CENTER: dep1:dep7;      # Center individual items
# Specify Regression Model
MODEL:
    # Define depression composite
    dep = dep1 + dep2 + dep3 + dep4 + dep5 + dep6 + dep7;

    focal:
        disab ~ 1@b0 dep@b1 male@b2 dep*male@b3;

    predictor:
        dep1:dep7 ~ 1 male;
```

Blimp Output for MODEL INFORMATION

```
CENTERED PREDICTORS
  Grand Mean Centered:    dep1 dep2 dep3 dep4 dep5 dep6 dep7

DEFINED VARIABLES
  [1]  dep = dep1+dep2+dep3+dep4+dep5+dep6+dep7

MODELS

 focal:
  [1]  disab ~ Intercept@b0 male@b2 dep@b1 dep*male@b3

 predictor:
  [2]  dep1 ~ Intercept dep2 dep3 dep4 dep5 dep6 dep7 male
  [3]  dep2 ~ Intercept dep3 dep4 dep5 dep6 dep7 male
  [4]  dep3 ~ Intercept dep4 dep5 dep6 dep7 male
  [5]  dep4 ~ Intercept dep5 dep6 dep7 male
  [6]  dep5 ~ Intercept dep6 dep7 male
  [7]  dep6 ~ Intercept dep7 male
  [8]  dep7 ~ Intercept male
```

Blimp PSRF Output

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 4 chains	Highest PSR	Parameter #
126 to 250	1.182	105
251 to 500	1.060	109
376 to 750	1.047	107
501 to 1000	1.032	109
⋮		
2001 to 4000	1.007	111
2126 to 4250	1.006	107
2251 to 4500	1.006	109
2376 to 4750	1.007	111
2501 to 5000	1.008	107

Worst $PSRF < 1.05$

Blimp Outcome Model Output for Focal Model

focal block:

Outcome Variable: *disab*

Parameters	Median	StdDev	2.5%	97.5%	PSR	N_Eff
Variances:						
Residual Var.	17.648	1.605	14.819	21.159	1.001	7433.659
Coefficients:						
Intercept	16.115	0.920	14.346	17.908	1.001	7802.323
male	2.950	1.399	0.168	5.669	1.000	8202.500
dep	0.412	0.060	0.294	0.528	1.001	7421.102
dep*male	−0.232	0.087	−0.404	−0.060	1.000	8348.711
Standardized Coefficients:						
male	0.310	0.144	0.018	0.588	1.000	8225.067
dep	0.546	0.070	0.398	0.673	1.001	8050.279
dep*male	−0.430	0.157	−0.729	−0.111	1.000	8380.377
Proportion Variance Explained						
by Coefficients	0.191	0.042	0.112	0.278	1.001	8191.891
by Residual Variation	0.809	0.042	0.722	0.888	1.001	8191.891

Conditional Effects Analysis (Simple Effects)

As a reminder, the conditional effects can be calculated using the following equation:

$$\begin{aligned} y_i &= \big(\text{simple intercept}\big) + \big(\text{simple slope}\big)\, x_i + e_i \\ &= (\beta_0 + \beta_2 m_i) + (\beta_1 + \beta_3 m_i)\, x_i + e_i \end{aligned}$$

Focal Model with Labels

```
focal:
    disab ~ 1@b0 dep@b1 male@b2 dep*male@b3;
```

PARAMETERS Command

```
# Conditional Effects Analysis (manually specify)
PARAMETERS:
    b0_female = b0;
    b1_female = b1;

    b0_male = b0 + b2;
    b1_male = b1 + b3;
```

Blimp Output for PARAMETERS Command

GENERATED PARAMETERS:

Summaries based on 10000 iterations using 4 chains.

Parameters	Median	StdDev	2.5%	97.5%	PSR	N_Eff
b0_female	16.115	0.920	14.346	17.908	1.001	7802.323
b1_female	0.412	0.060	0.294	0.528	1.001	7421.102
b0_male	19.076	1.045	17.017	21.096	1.000	8811.626
b1_male	0.180	0.063	0.058	0.305	1.000	8501.354

Blimp Software

- General-purpose Bayesian estimation for regression and path models.
- Single and multilevel models
- Allows for latent variables, incomplete predictors and outcomes
- Interactive and nonlinear effects
- Nonnormal data
- And more!

Available at

`https://www.appliedmissingdata.com/blimp`