Data Analysis and Statistical Software

Analyzing data with missing values using multiple imputation

Meghan Cain | September 29, 2020

You can download the datasets and do-file here:
http://tinyurl.com/mi-web-2020

Missing Data Mechanisms

- Missing Completely At Random (MCAR):
  - The missingness is unrelated to any of the variables in the model
  - Missing values are a simple random sample of all data values
- Missing At Random (MAR):
  - The missingness is related to the observed variables
  - Missing values are a simple random sample of all data values conditional on the observed data
- Missing Not At Random (MNAR):
  - The missingness is related to the unobserved variables
  - Missing values are not a simple random sample of all data values

Missing Data Analysis

- Listwise deletion is inefficient and can result in bias under MAR.
- Single imputation methods underestimate the standard errors and can result in bias under MCAR and MAR.
- Multiple imputation (MI) is a “state-of-the-art” missing data approach that results in efficient, valid statistical inference for data that are either MCAR and MAR.
- MI is a simulation-based approach for analyzing incomplete data that involves filling in missing responses multiple times.
- MI is often regarded as the most flexible missing data approach.
  - It can be used with virtually any analysis model
  - The imputation model can include auxiliary variables
  - Separate people can perform the imputation and the analysis
  - One set of imputations can be used for several analysis models

Three Steps of MI

Original Data

x1	x2	x3
4	2	
6		5
9	1	4

Three Steps of MI

Original Data

x1	x2	x3
4	2	
6		5
9	1	4

1. Imputation

$m = 1$

x1	x2	x3
4	2	3
6	1	5
9	1	4

$m = 2$

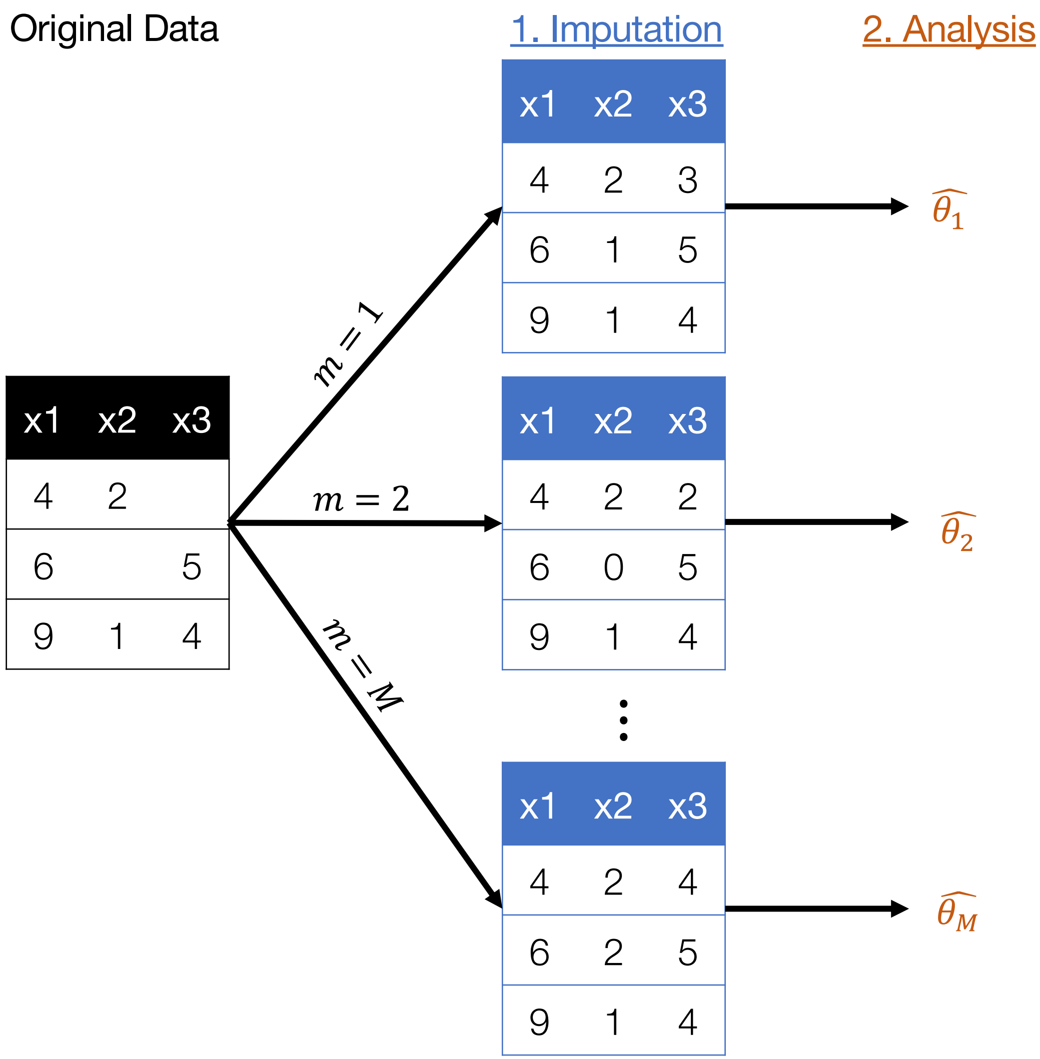
x1	x2	x3
4	2	2
6	0	5
9	1	4

⋮

$m = M$

x1	x2	x3
4	2	4
6	2	5
9	1	4

Three Steps of MI



Original Data

x1	x2	x3
4	2	
6		5
9	1	4

1. Imputation ($m = 1$)

x1	x2	x3
4	2	3
6	1	5
9	1	4

($m = 2$)

x1	x2	x3
4	2	2
6	0	5
9	1	4

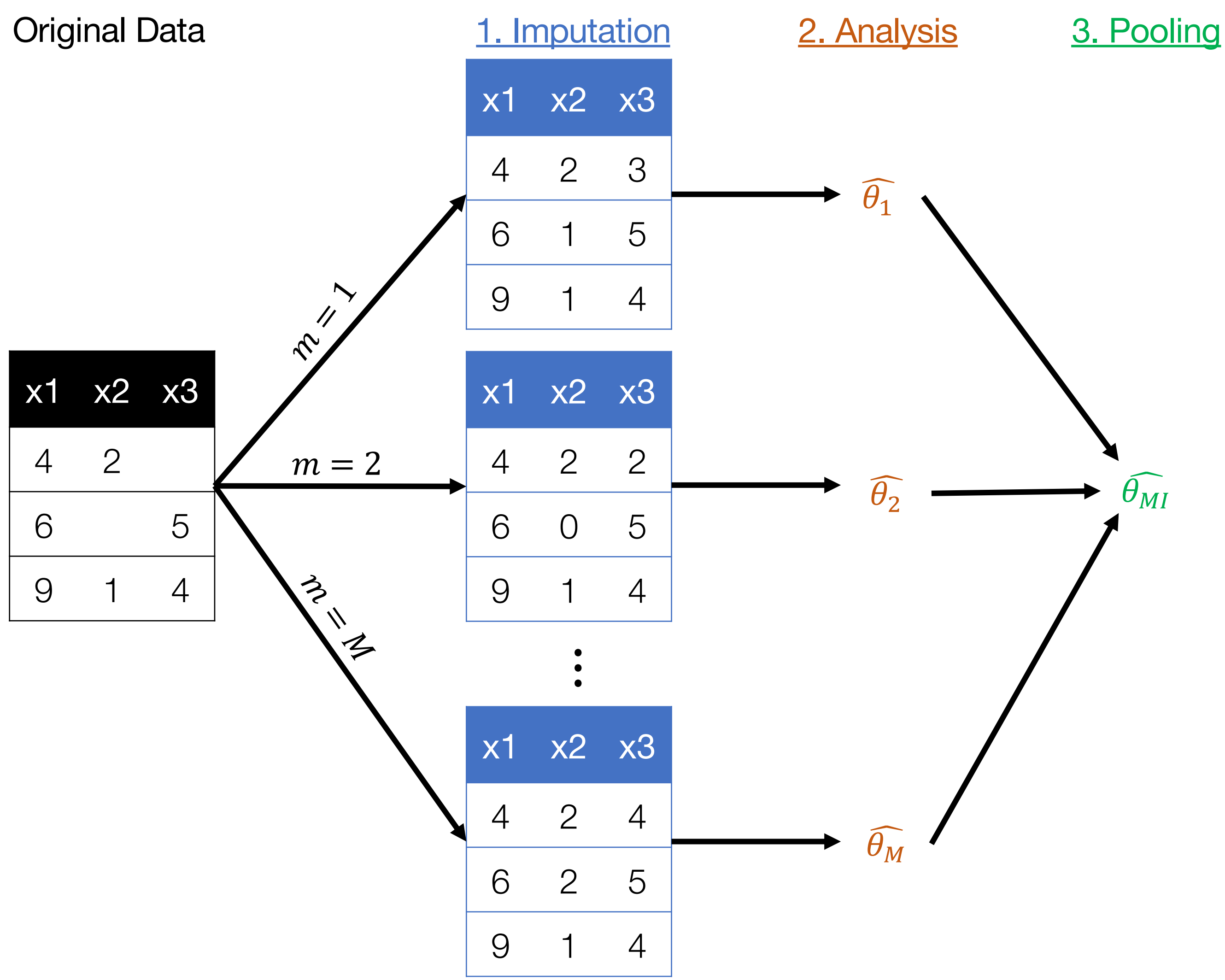
⋮

($m = M$)

x1	x2	x3
4	2	4
6	2	5
9	1	4

2. Analysis: $\widehat{\theta_1}$, $\widehat{\theta_2}$, $\widehat{\theta_M}$

Three Steps of MI



An Example

```
. use heart, clear

. summarize

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
      attack |       154    .4480519    .4989166          0          1
      smokes |       154    .4155844    .4944304          0          1
         age |       154    56.48829    11.73051   20.73613   87.14446
         bmi |       132    25.24136    4.027137   17.22643   38.24214
      hsgrad |       154    .7532468    .4325285          0          1
-------------+--------------------------------------------------------
      female |       154    .2467532    .4325285          0          1
```

Complete-case analysis

```
. logit attack smokes age bmi hsgrad female, or nolog

Logistic regression                             Number of obs     =        132
                                                LR chi2(5)        =      24.03
                                                Prob > chi2       =     0.0002
Log likelihood =  -79.34221                     Pseudo R2         =     0.1315

------------------------------------------------------------------------------
      attack | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      smokes |   4.683533   1.872631     3.86   0.000      2.13912    10.25444
         age |   1.026456   .0174929     1.53   0.125     .9927368     1.06132
         bmi |   1.119625   .0559881     2.26   0.024     1.015096    1.234917
      hsgrad |    1.49904   .6664762     0.91   0.363     .6271443    3.583102
      female |   1.252987    .567297     0.50   0.618     .5158935    3.043217
       _cons |   .0044788   .0081094    -2.99   0.003     .0001288    .1557224
------------------------------------------------------------------------------
Note: _cons estimates baseline odds.
```

Multiple Imputation Analysis

```
. mi impute regress bmi attack smokes age hsgrad female, add(20) rseed(298127)

. mi estimate, or: logit attack smokes age bmi hsgrad female

Multiple-imputation estimates                   Imputations       =         20
Logistic regression                             Number of obs     =        154
                                                Average RVI       =     0.0647
                                                Largest FMI       =     0.2576
DF adjustment:   Large sample                   DF:     min       =     297.64
                                                        avg       = 100,526.42
                                                        max       = 429,095.88
Model F test:       Equal FMI                   F(   5,17291.9)   =       3.43
Within VCE type:          OIM                   Prob > F          =     0.0042

------------------------------------------------------------------------------
      attack | Odds Ratio   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      smokes |   3.355972   1.209766     3.36   0.001     1.655651    6.802487
         age |   1.036412   .0159889     2.32   0.020     1.005543    1.068228
         bmi |   1.107417   .0568467     1.99   0.048      1.00101    1.225135
      hsgrad |   1.188189   .4837912     0.42   0.672     .5349293    2.639216
      female |    .908759   .3806569    -0.23   0.819      .399848    2.065392
       _cons |    .004372   .0077669    -3.06   0.002     .0001335    .1431512
------------------------------------------------------------------------------
Note: _cons estimates baseline odds.
```

`mi` suite

- Examine: `misstable`, `mi describe`
- Setup: `mi set`, `mi register`
- Impute: `mi impute`
- Analysis and Pooling: `mi estimate`
- Test: `mi test`, `mi testtransform`
- Predict: `mi predict`, `mi predictnl`
- Import: `mi import`
- Manage: `mi merge`, `mi reshape`, `mi xeq`, and more

Steps of MI

1. Setup
2. Imputation
3. Analysis
4. Pooling
5. Postestimation

- Importing
- Data Management

Steps of MI

1. Setup
2. Imputation
3. Analysis
4. Pooling
5. Postestimation

- Importing
- Data Management

Setup

- Choose an `mi` style (how imputations are stored)
  - `wide`
  - `mlong`
  - `flong`
  - `flongsep`

mi Styles

wide

x	z	_mi_miss	_1_x	_2_x
5	21	0	5	5
.	26	1	4.5	4
3	30	0	3	3

mlong

x	z	_mi_miss	_mi_id	_mi_m
5	21	0	1	0
.	26	1	2	0
3	30	0	3	0
4.5	26	.	2	1
4	26	.	2	2

mi Styles

flong

x	z	_mi_miss	_mi_id	_mi_m
5	21	0	1	0
.	26	1	2	0
3	30	0	3	0
5	21	.	1	1
4.5	26	.	2	1
3	30	.	3	1
5	21	.	1	2
4	26	.	2	2
3	30	.	3	2

flongsep

x	z	_mi_miss	_mi_id
5	21	0	1
4.5	26	1	2
3	30	0	3

_1_dat.dta

x	z	_mi_miss	_mi_id
5	21	0	1
4	26	1	2
3	30	0	3

_2_dat.dta

Setup

- Choose an `mi` style (how imputations are stored)

```
. mi set wide
```

Setup

- Choose an `mi` style (how imputations are stored)

```
. mi set wide
```

- Register variables

```
. mi register imputed bmi
. mi register regular attack smokes age hsgrad female
```

Steps of MI

1. Setup
2. Imputation
3. Analysis
4. Pooling
5. Postestimation

- Importing
- Data Management

Imputation: Models

`mi impute` *imputation_ method*

Pattern	Type	Imputation Method
Univariate	Continuous	`regress`, `pmm`, `truncreg`, `intreg`
	Binary	`logit`
	Categorical	`ologit`, `mlogit`
	Count	`poisson`, `nbreg`
Monotone	Mixture	`monotone`
Arbitrary	Continuous	`mvn`
	Mixture	`chained`

Imputation: `regress`

- `mi impute regress` assumes there is one normally-distributed variable (conditionally on complete predictors) with missing observations.
- Use a linear regression model to fill in missing observations, adding random variability each time to create M unique imputations.
- To demonstrate, we will partition the dataset into two groups, $\boldsymbol{X} = \{\boldsymbol{X_{obs}}, \boldsymbol{X_{mis}}\}$, where $\boldsymbol{X_{obs}}$ contain the complete observations and $\boldsymbol{X_{mis}}$ contain the observations with missing responses.

Imputation: `regress`

$\texttt{bmi}_{\text{obs}} = \beta_0 + \beta_1\texttt{attack}_{\text{obs}} + \beta_2\texttt{smokes}_{\text{obs}} + \beta_3\texttt{age}_{\text{obs}} + \beta_4\texttt{hsgrad}_{\text{obs}} + \beta_5\texttt{female}_{\text{obs}} + \varepsilon$

where $var(\varepsilon) = \sigma^2$

Imputation: `regress`

$\texttt{bmi}_{\texttt{obs}} = \beta_0 + \beta_1 \texttt{attack}_{\texttt{obs}} + \beta_2 \texttt{smokes}_{\texttt{obs}} + \beta_3 \texttt{age}_{\texttt{obs}} + \beta_4 \texttt{hsgrad}_{\texttt{obs}} + \beta_5 \texttt{female}_{\texttt{obs}} + \varepsilon$

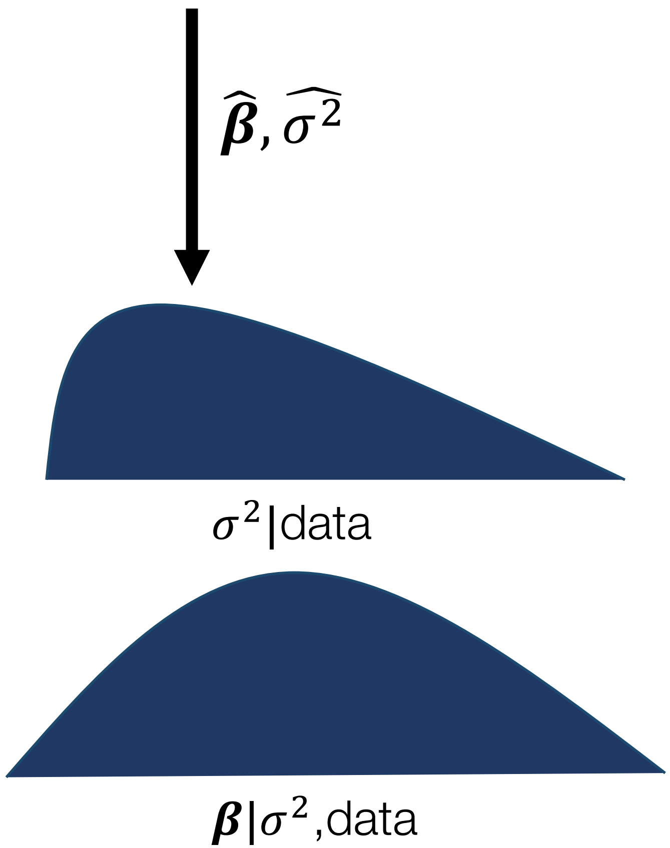
where $var(\varepsilon) = \sigma^2$

$\downarrow \; \widehat{\boldsymbol{\beta}}, \widehat{\sigma^2}$

Imputation: `regress`

$\texttt{bmi}_{\text{obs}} = \beta_0 + \beta_1 \texttt{attack}_{\text{obs}} + \beta_2 \texttt{smokes}_{\text{obs}} + \beta_3 \texttt{age}_{\text{obs}} + \beta_4 \texttt{hsgrad}_{\text{obs}} + \beta_5 \texttt{female}_{\text{obs}} + \varepsilon$

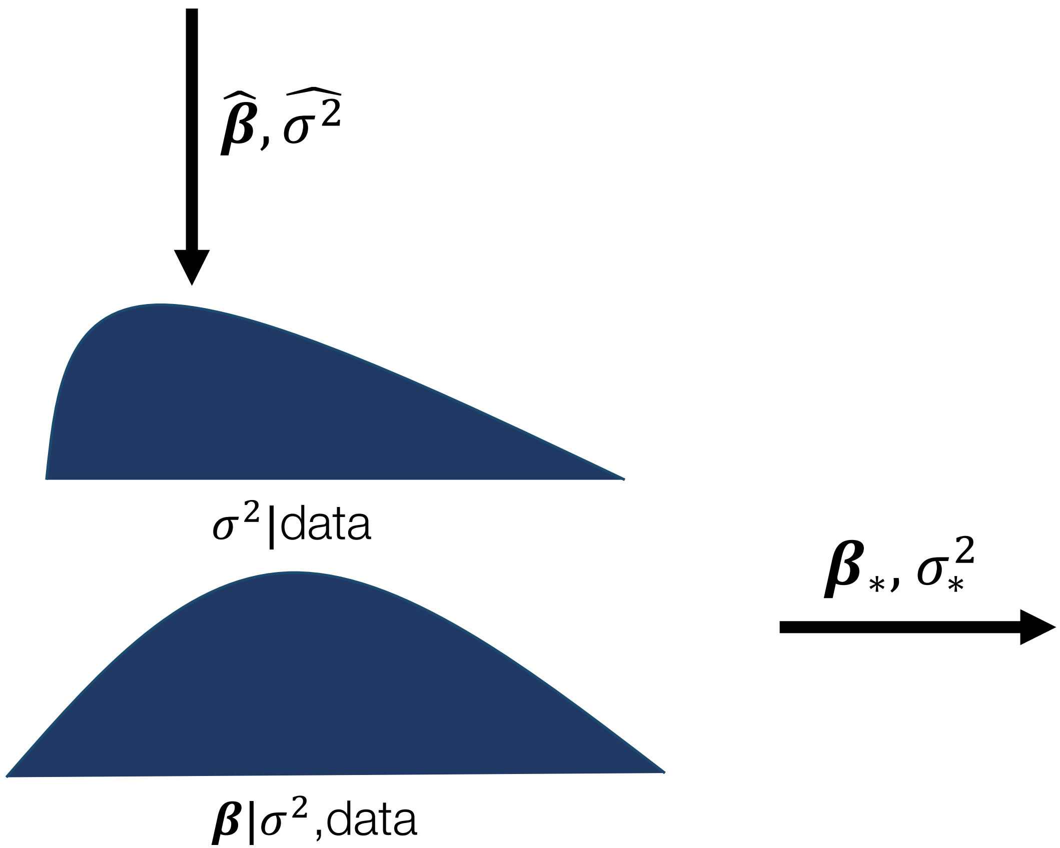
where $var(\varepsilon) = \sigma^2$



Imputation: `regress`

$\texttt{bmi}_{\text{obs}} = \beta_0 + \beta_1\texttt{attack}_{\text{obs}} + \beta_2\texttt{smokes}_{\text{obs}} + \beta_3\texttt{age}_{\text{obs}} + \beta_4\texttt{hsgrad}_{\text{obs}} + \beta_5\texttt{female}_{\text{obs}} + \varepsilon$

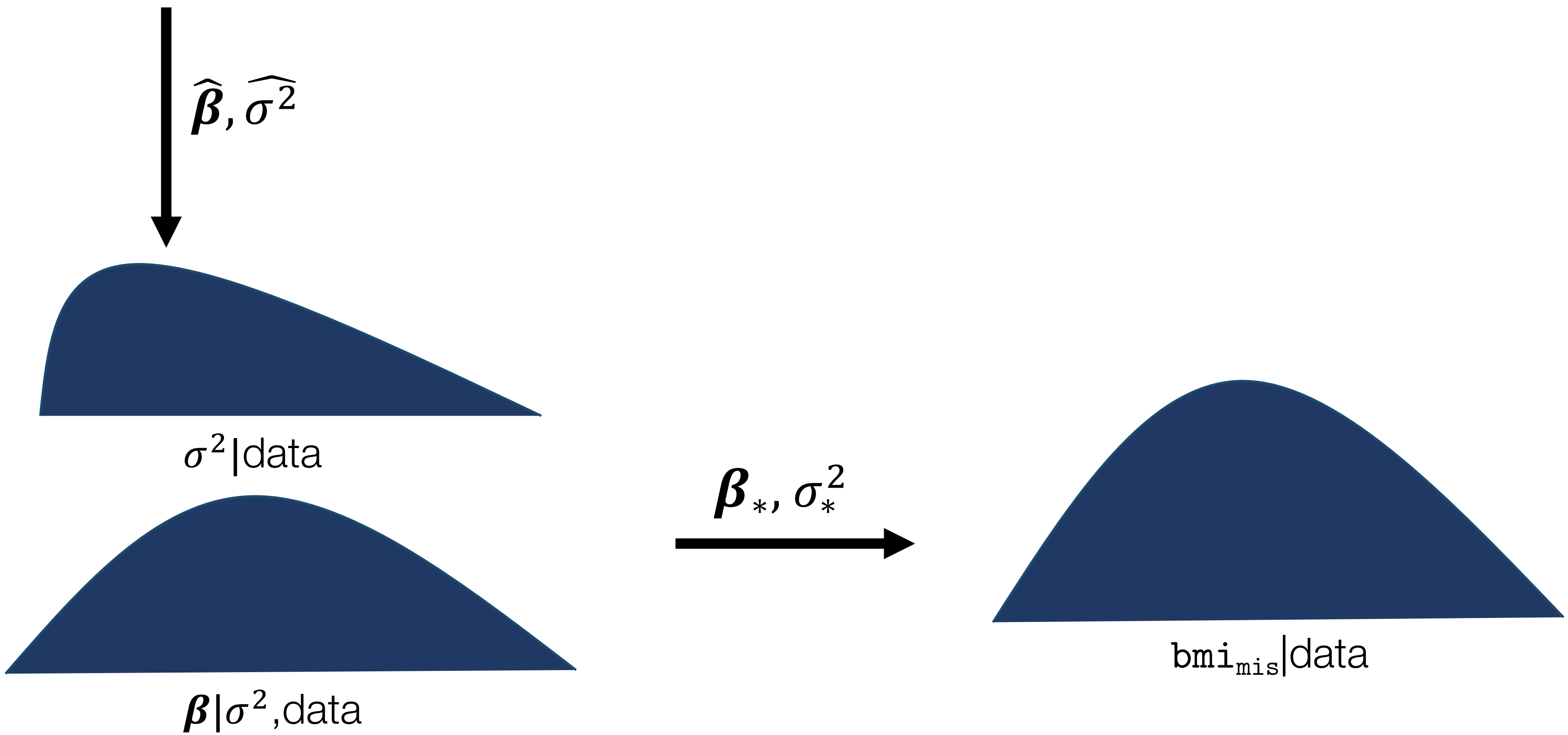
where $var(\varepsilon) = \sigma^2$



Imputation: `regress`

$\texttt{bmi}_{\texttt{obs}} = \beta_0 + \beta_1 \texttt{attack}_{\texttt{obs}} + \beta_2 \texttt{smokes}_{\texttt{obs}} + \beta_3 \texttt{age}_{\texttt{obs}} + \beta_4 \texttt{hsgrad}_{\texttt{obs}} + \beta_5 \texttt{female}_{\texttt{obs}} + \varepsilon$

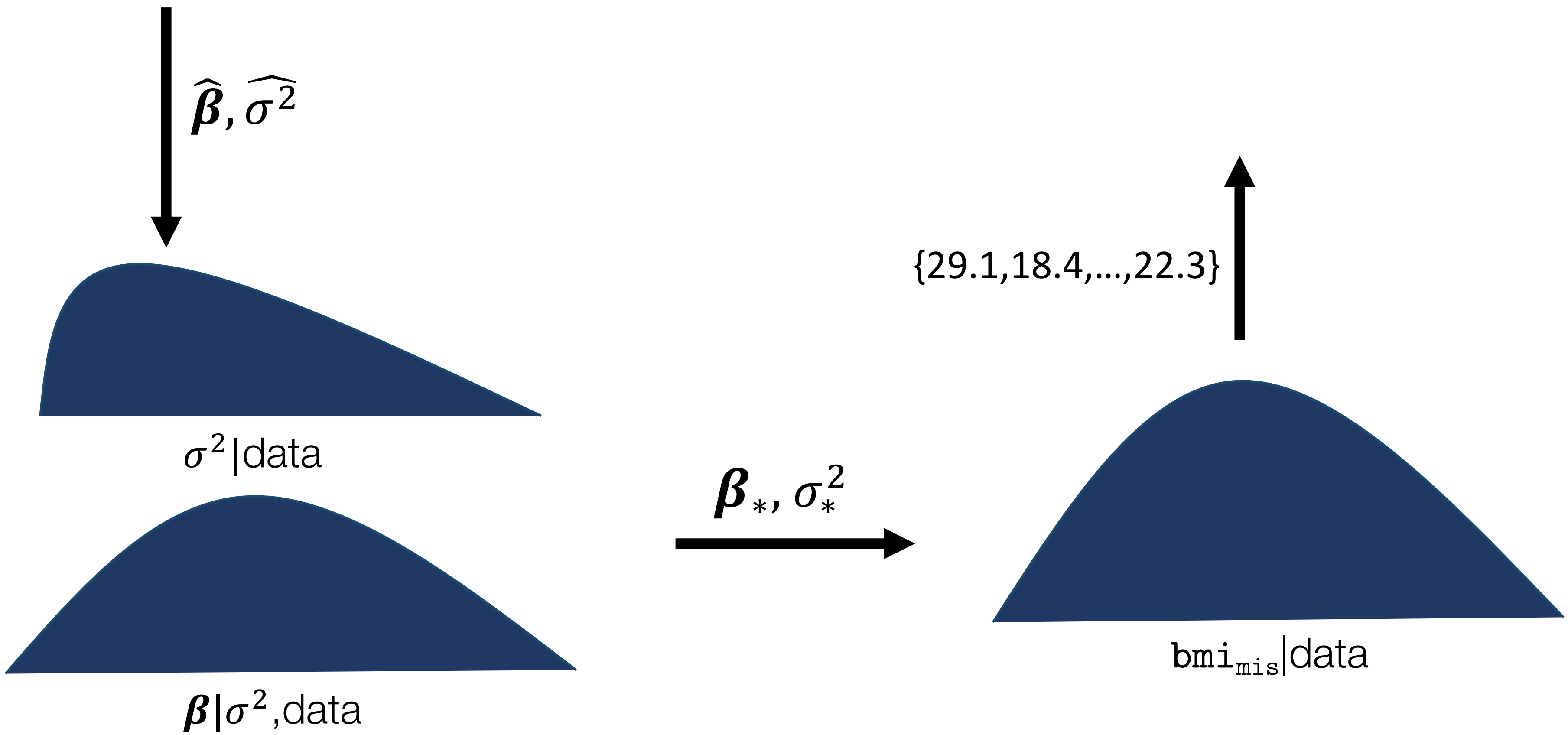
where $var(\varepsilon) = \sigma^2$



Imputation: `regress`

$\texttt{bmi}_{\texttt{obs}} = \beta_0 + \beta_1 \texttt{attack}_{\texttt{obs}} + \beta_2 \texttt{smokes}_{\texttt{obs}} + \beta_3 \texttt{age}_{\texttt{obs}} + \beta_4 \texttt{hsgrad}_{\texttt{obs}} + \beta_5 \texttt{female}_{\texttt{obs}} + \varepsilon$

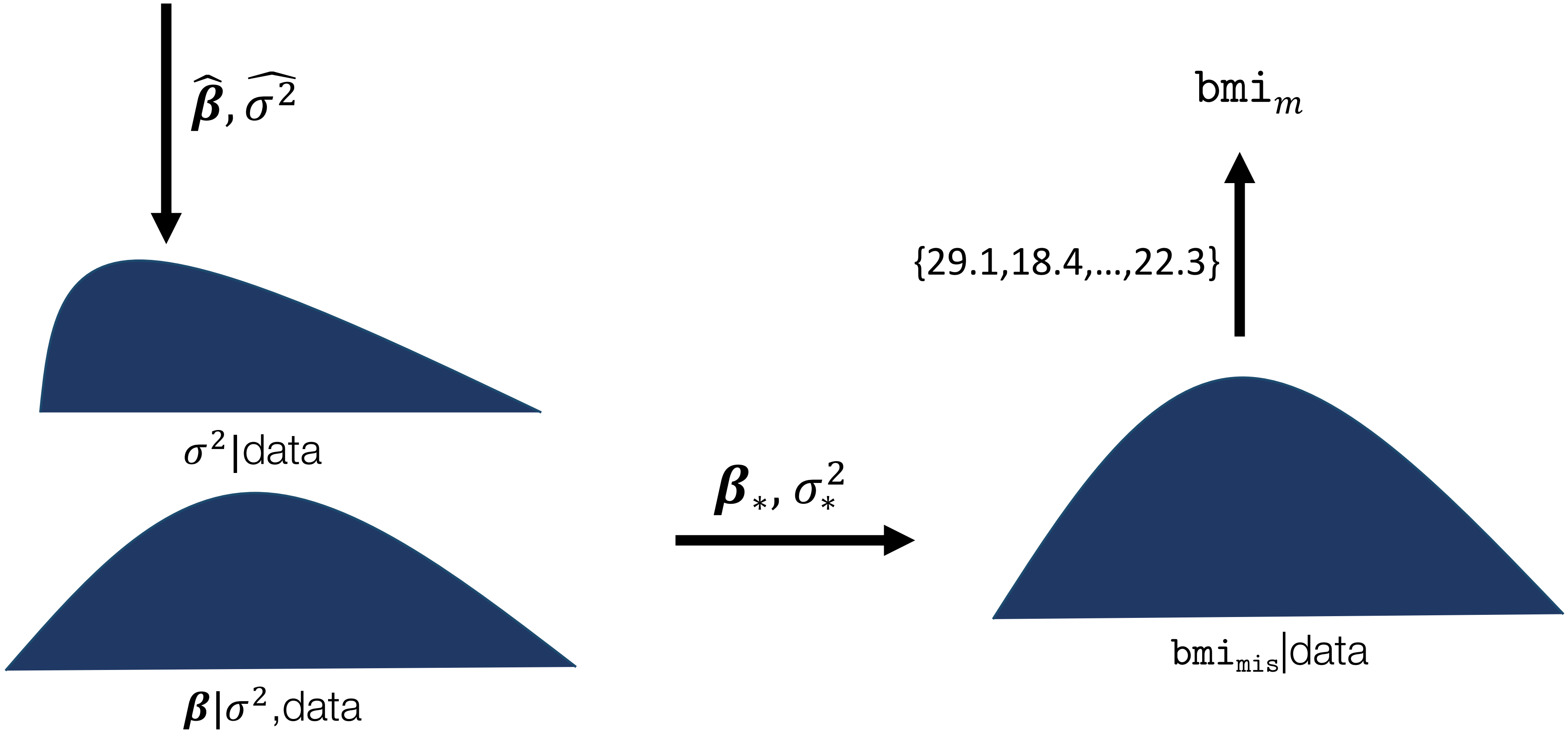
where $var(\varepsilon) = \sigma^2$



Imputation: `regress`

$\texttt{bmi}_{\texttt{obs}} = \beta_0 + \beta_1 \texttt{attack}_{\texttt{obs}} + \beta_2 \texttt{smokes}_{\texttt{obs}} + \beta_3 \texttt{age}_{\texttt{obs}} + \beta_4 \texttt{hsgrad}_{\texttt{obs}} + \beta_5 \texttt{female}_{\texttt{obs}} + \varepsilon$

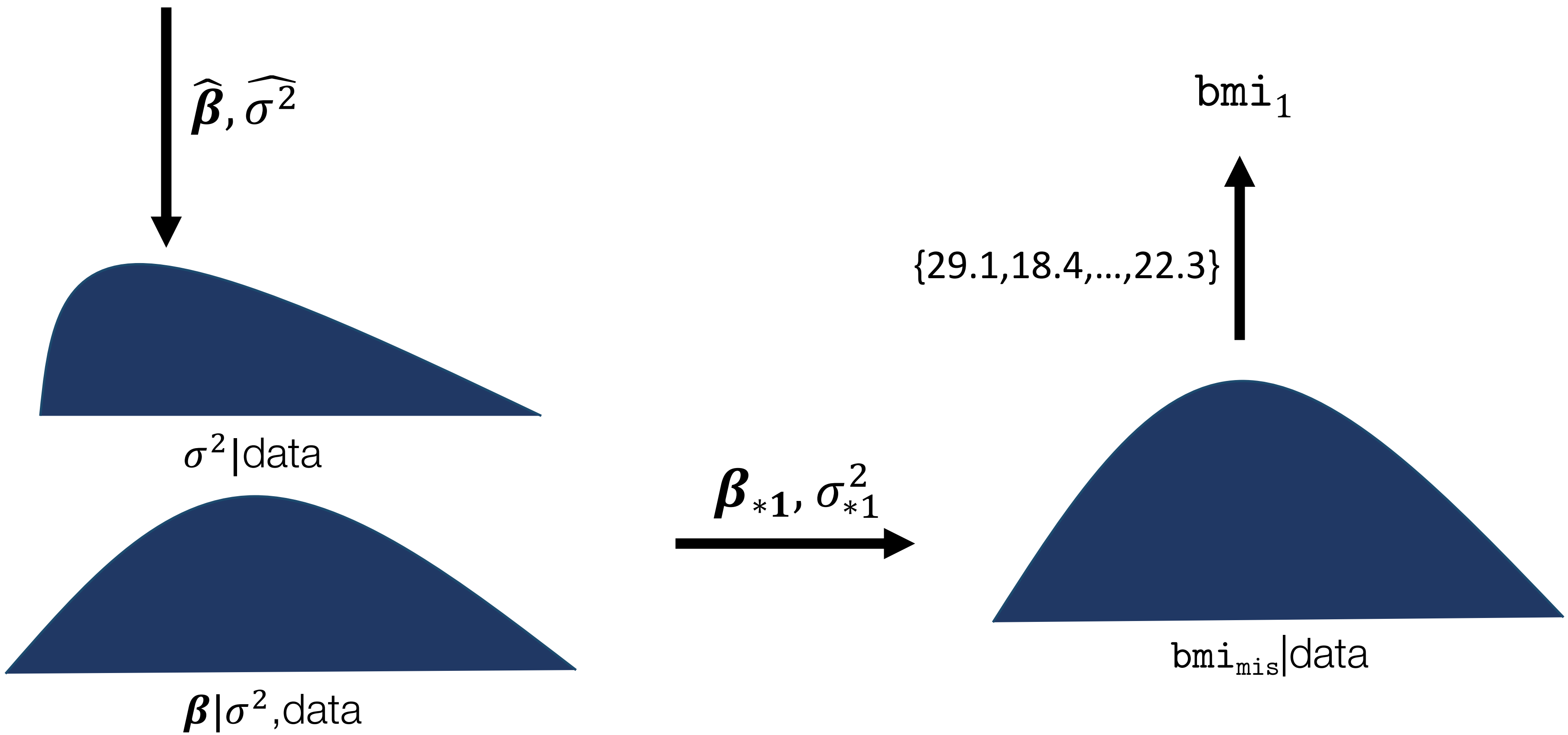
where $var(\varepsilon) = \sigma^2$



Imputation: `regress`

$\texttt{bmi}_{\texttt{obs}} = \beta_0 + \beta_1 \texttt{attack}_{\texttt{obs}} + \beta_2 \texttt{smokes}_{\texttt{obs}} + \beta_3 \texttt{age}_{\texttt{obs}} + \beta_4 \texttt{hsgrad}_{\texttt{obs}} + \beta_5 \texttt{female}_{\texttt{obs}} + \varepsilon$

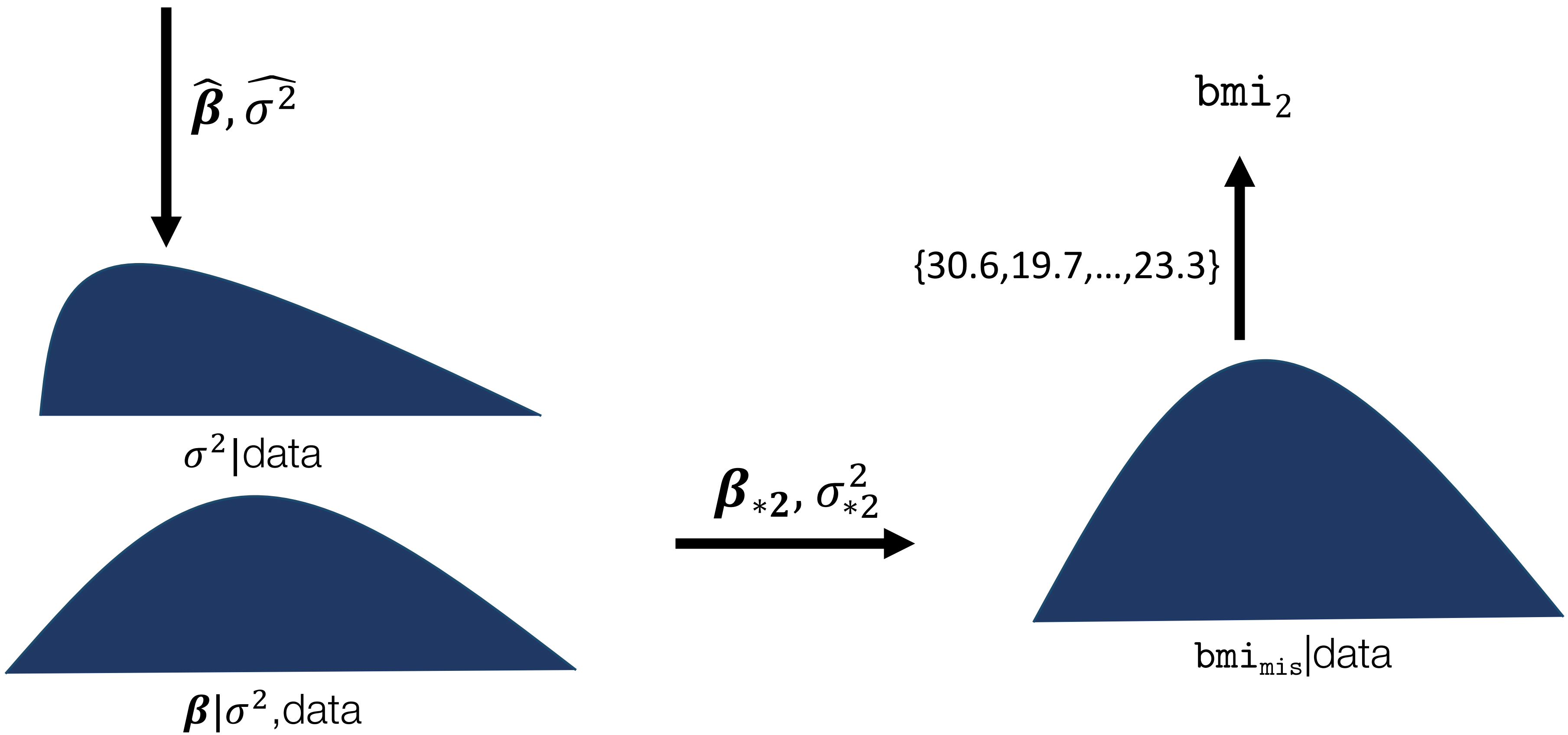
where $var(\varepsilon) = \sigma^2$



Imputation: `regress`

$\texttt{bmi}_{\texttt{obs}} = \beta_0 + \beta_1\texttt{attack}_{\texttt{obs}} + \beta_2\texttt{smokes}_{\texttt{obs}} + \beta_3\texttt{age}_{\texttt{obs}} + \beta_4\texttt{hsgrad}_{\texttt{obs}} + \beta_5\texttt{female}_{\texttt{obs}} + \varepsilon$

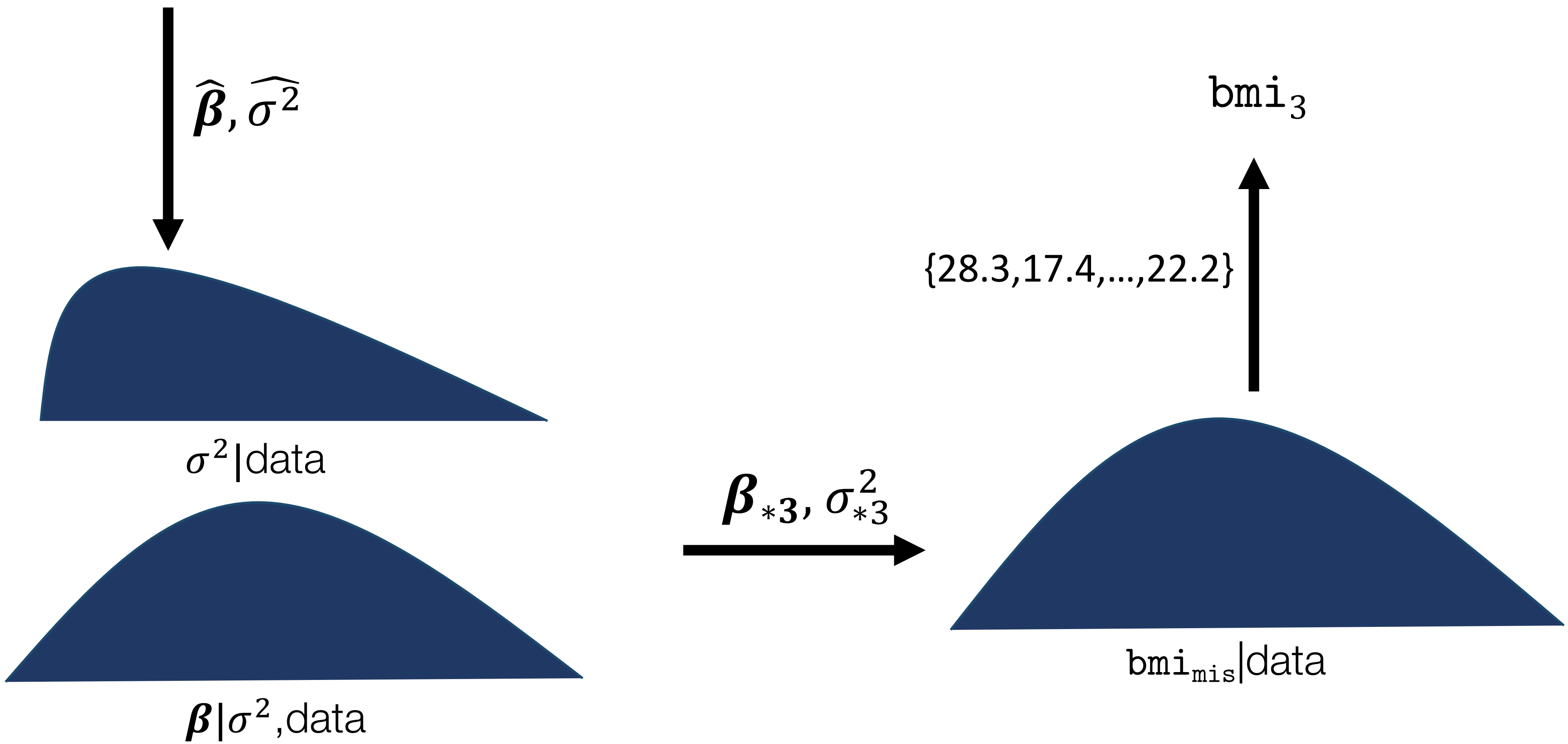
where $var(\varepsilon) = \sigma^2$



Imputation: `regress`

$\texttt{bmi}_{\texttt{obs}} = \beta_0 + \beta_1 \texttt{attack}_{\texttt{obs}} + \beta_2 \texttt{smokes}_{\texttt{obs}} + \beta_3 \texttt{age}_{\texttt{obs}} + \beta_4 \texttt{hsgrad}_{\texttt{obs}} + \beta_5 \texttt{female}_{\texttt{obs}} + \varepsilon$

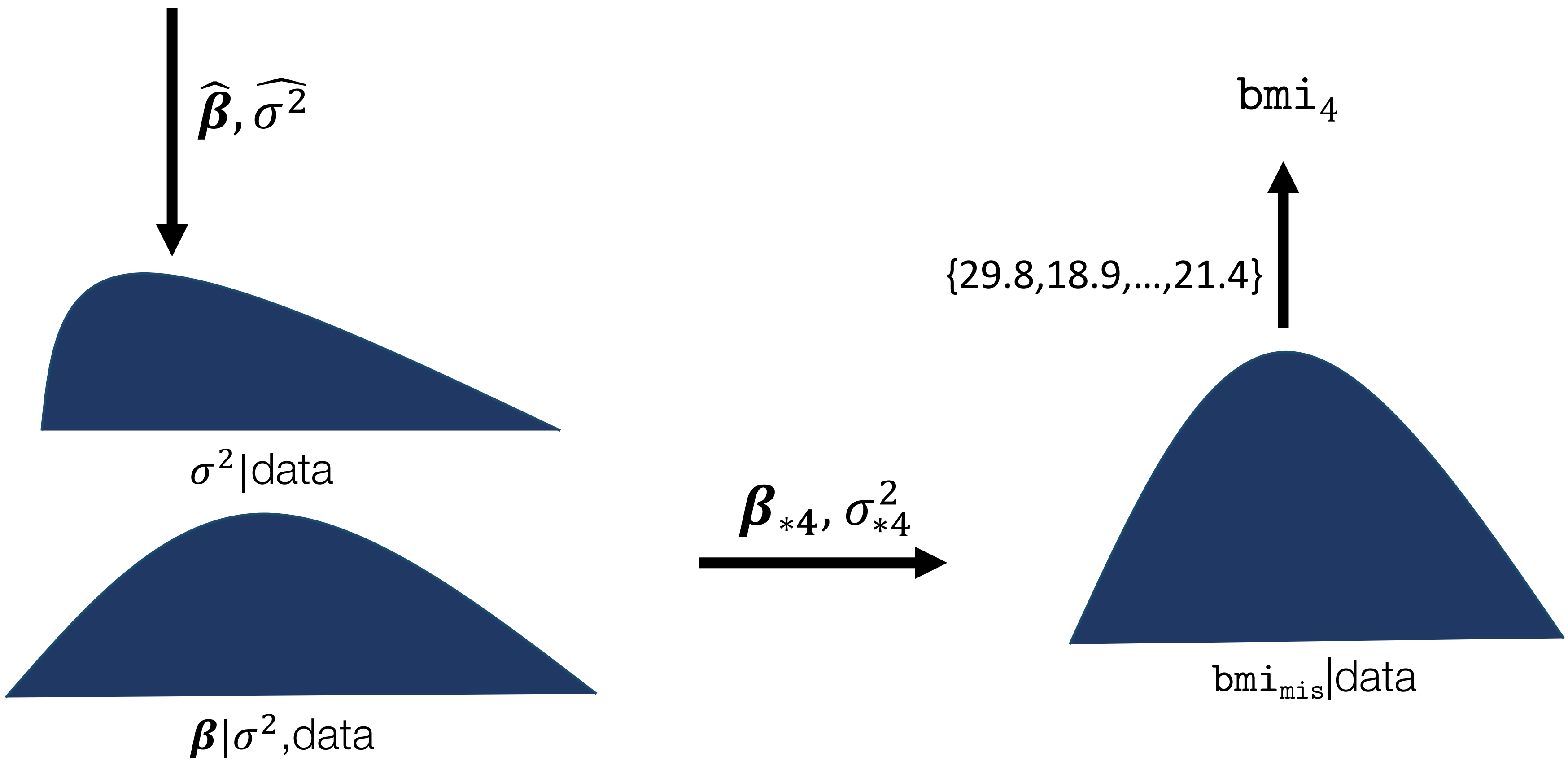
where $var(\varepsilon) = \sigma^2$



Imputation: `regress`

$\texttt{bmi}_{\texttt{obs}} = \beta_0 + \beta_1 \texttt{attack}_{\texttt{obs}} + \beta_2 \texttt{smokes}_{\texttt{obs}} + \beta_3 \texttt{age}_{\texttt{obs}} + \beta_4 \texttt{hsgrad}_{\texttt{obs}} + \beta_5 \texttt{female}_{\texttt{obs}} + \varepsilon$

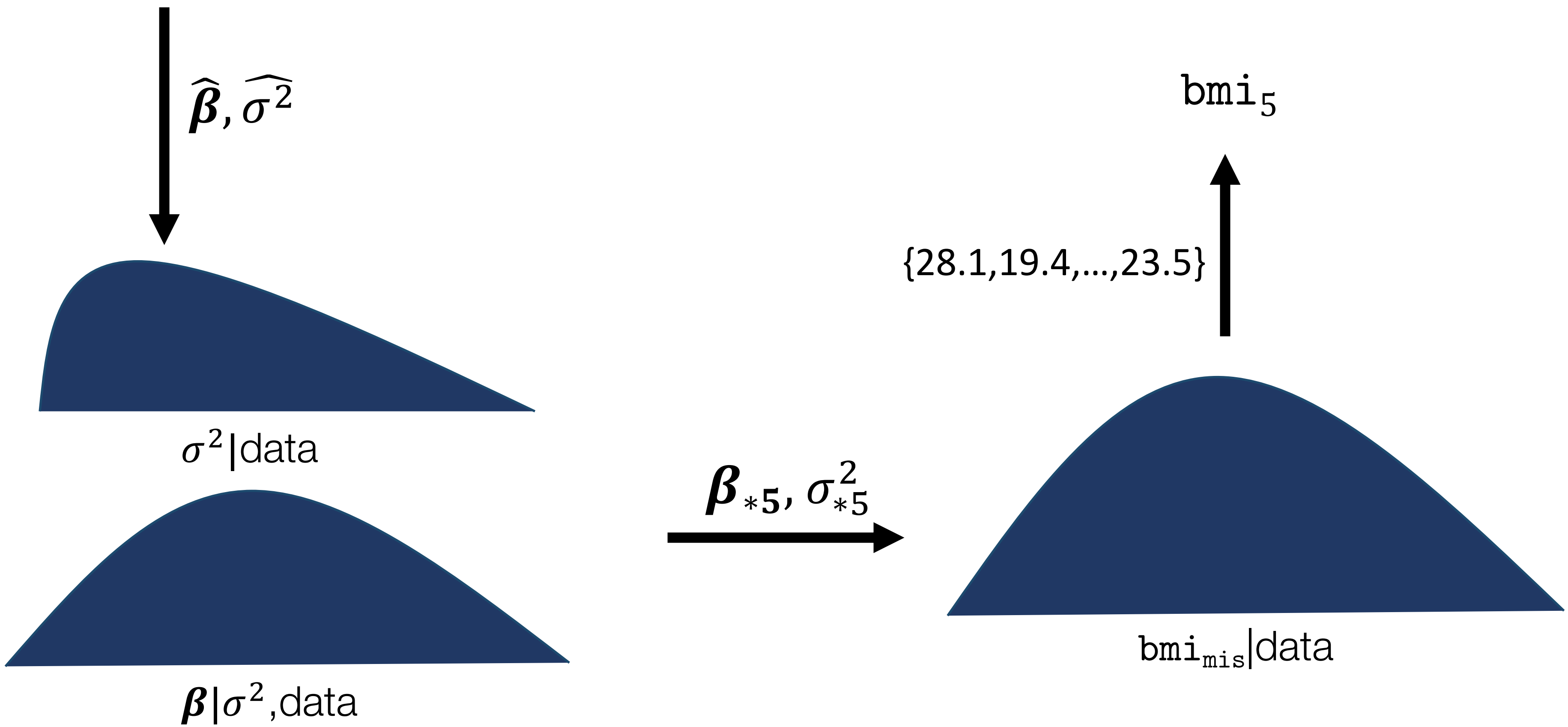
where $var(\varepsilon) = \sigma^2$



Imputation: `regress`

$\texttt{bmi}_{\texttt{obs}} = \beta_0 + \beta_1\texttt{attack}_{\texttt{obs}} + \beta_2\texttt{smokes}_{\texttt{obs}} + \beta_3\texttt{age}_{\texttt{obs}} + \beta_4\texttt{hsgrad}_{\texttt{obs}} + \beta_5\texttt{female}_{\texttt{obs}} + \varepsilon$

where $var(\varepsilon) = \sigma^2$



Imputation: `regress`

$\texttt{bmi}_{\text{obs}} = \beta_0 + \beta_1\texttt{attack}_{\text{obs}} + \beta_2\texttt{smokes}_{\text{obs}} + \beta_3\texttt{age}_{\text{obs}} + \beta_4\texttt{hsgrad}_{\text{obs}} + \beta_5\texttt{female}_{\text{obs}} + \varepsilon$

where $var(\varepsilon) = \sigma^2$

```
. mi impute regress bmi attack smokes age hsgrad female, add(20) rseed(298127)

Univariate imputation                          Imputations =       20
Linear regression                                    added =       20
Imputed: m=1 through m=20                          updated =        0
```

Variable	Observations per m			
	Complete	Incomplete	Imputed	Total
bmi	132	22	22	154

(complete + incomplete = total; imputed is the minimum across *m* of the number of filled-in observations.)

Imputation: regress

```
. list bmi _1_bmi _2_bmi _3_bmi _4_bmi _5_bmi if _n<6
```

	bmi	_1_bmi	_2_bmi	_3_bmi	_4_bmi	_5_bmi
1.	21.11455	21.11455	21.11455	21.11455	21.11455	21.11455
2.	24.8684	24.8684	24.8684	24.8684	24.8684	24.8684
3.	30.50274	30.50274	30.50274	30.50274	30.50274	30.50274
4.	.	29.88588	21.41766	24.19195	19.40182	26.64958
5.	22.52744	22.52744	22.52744	22.52744	22.52744	22.52744

Multivariate Multiple Imputation

- Let's consider the same dataset, with some additional missing observations for `smokes`

```
. use heart2_miset, clear

. mi describe

  Style:  wide
          last mi update 16sep2020 16:47:25, approximately 46 hours ago

  Obs.:   complete          120
          incomplete         34  (M = 0 imputations)
          ---------------------
          total             154

  Vars.:  imputed:  2; bmi(22) smokes(16)

          passive:  0

          regular:  4; attack age hsgrad female

          system:   1; _mi_miss

         (there are no unregistered variables)
```

Imputation: `chained`

- Multiple imputation using chained equations (ICE) is performed by `mi impute chained`.
- The pattern of missing data can be arbitrary.
- Variables are imputed iteratively using conditional univariate imputation models

$$P(\texttt{smokes}^{t} | \texttt{bmi}^{t-1}, \texttt{attack}, \texttt{age}, \texttt{hsgrad}, \texttt{female}, \theta)$$

$$P(\texttt{bmi}^{t} | \texttt{smokes}^{t}, \texttt{attack}, \texttt{age}, \texttt{hsgrad}, \texttt{female}, \theta)$$

Imputation: `chained`

```
. mi impute chained (regress) bmi (logit) smokes = attack age hsgrad female, ///
> add(20) rseed(298127)

Conditional models:
           smokes: logit smokes bmi attack age hsgrad female
              bmi: regress bmi i.smokes attack age hsgrad female

Performing chained iterations ...

Multivariate imputation                     Imputations =       20
Chained equations                                 added =       20
Imputed: m=1 through m=20                       updated =        0

Initialization: monotone                     Iterations =      200
                                                burn-in =       10

              bmi: linear regression
           smokes: logistic regression
```

Variable	Observations per m			
	Complete	Incomplete	Imputed	Total
bmi	132	22	22	154
smokes	138	16	16	154

(complete + incomplete = total; imputed is the minimum across *m* of the number of filled-in observations.)

Imputation: `chained`

```
. mi impute chained (regress) bmi (logit) smokes = attack age hsgrad female, ///
> add(20) rseed(298127)

Conditional models:
           smokes: logit smokes bmi attack age hsgrad female
              bmi: regress bmi i.smokes attack age hsgrad female

Performing chained iterations ...

Multivariate imputation                     Imputations =       20
Chained equations                                 added =       20
Imputed: m=1 through m=20                       updated =        0

Initialization: monotone                     Iterations =      200
                                                burn-in =       10

              bmi: linear regression
           smokes: logistic regression
```

Variable	Complete	Incomplete	Imputed	Total
	Observations per *m*			
bmi	132	22	22	154
smokes	138	16	16	154

```
(complete + incomplete = total; imputed is the minimum across m
 of the number of filled-in observations.)
```

Imputation: `chained`

```
. mi impute chained (regress) bmi (logit) smokes = attack age hsgrad female, ///
> add(20) rseed(298127)

Conditional models:
           smokes: logit smokes bmi attack age hsgrad female
              bmi: regress bmi i.smokes attack age hsgrad female

Performing chained iterations ...

Multivariate imputation                    Imputations =       20
Chained equations                                added =       20
Imputed: m=1 through m=20                      updated =        0

Initialization: monotone                    Iterations =      200
                                               burn-in =       10

              bmi: linear regression
           smokes: logistic regression
```

Variable	Observations per m			
	Complete	Incomplete	Imputed	Total
bmi	132	22	22	154
smokes	138	16	16	154

(complete + incomplete = total; imputed is the minimum across *m* of the number of filled-in observations.)

Steps of MI

1. Setup
2. Imputation
3. Analysis
4. Pooling
5. Postestimation

- Importing
- Data Management

Analysis Models

`mi estimate:` *estimation_ command*

- `regress` Linear regression
- `logit` Logistic regression
- `poisson` Poisson regression
- `stcox` Cox proportional hazards model
- `glm` Generalized linear models
- `xtreg` Fixed- and random-effects and PA linear models
- `mixed` Multilevel mixed-effects linear regression
- `svy:` Estimation commands for survey data

For a full list type `help mi estimate`

Estimate

```
. mi estimate, or: logit attack smokes age bmi hsgrad female

Multiple-imputation estimates                  Imputations       =         20
Logistic regression                            Number of obs     =        154
                                               Average RVI       =     0.0831
                                               Largest FMI       =     0.2301
DF adjustment:   Large sample                  DF:     min       =     372.02
                                                       avg       =  45,931.24
                                                       max       = 123,886.91
Model F test:       Equal FMI                  F(   5,11115.9)   =       3.46
Within VCE type:          OIM                  Prob > F          =     0.0039

------------------------------------------------------------------------------
      attack | Odds Ratio   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      smokes |   3.427856   1.326603     3.18   0.001     1.604092    7.325141
         age |    1.03589   .0160632     2.27   0.023      1.00488    1.067857
         bmi |    1.11175   .0567782     2.07   0.039     1.005527    1.229195
      hsgrad |   1.214893   .4965427     0.48   0.634      .545293    2.706739
      female |    .904324    .378137    -0.24   0.810     .3984704    2.052353
       _cons |   .0040567   .0070491    -3.17   0.002     .0001341    .1227637
------------------------------------------------------------------------------
Note: _cons estimates baseline odds.
```

Estimate

```
. mi estimate, vartable nocitable

Multiple-imputation estimates                     Imputations       =         20
Logistic regression

Variance information
```

	Imputation variance Within	Between	Total	RVI	FMI	Relative efficiency
smokes	.129823	.019001	.149774	.153683	.134826	.993304
age	.000237	2.8e-06	.00024	.012539	.0124	.99938
bmi	.002019	.000561	.002608	.291978	.230121	.988625
hsgrad	.163308	.003561	.167046	.022893	.022433	.99888
female	.17256	.002175	.174844	.013236	.013081	.999346
_cons	2.59897	.400332	3.01932	.161736	.14097	.993001

Estimate

```
. mi estimate, mcerror noheader: logit attack smokes age bmi hsgrad female
```

attack	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
smokes	1.231935	.3870065	3.18	0.001	.4725578	1.991312
	.0308233	.0080164	0.12	0.001	.0385576	.0306408
age	.0352609	.0155066	2.27	0.023	.0048682	.0656536
	.0003766	.0000595	0.02	0.001	.0003381	.0004433
bmi	.1059358	.051071	2.07	0.039	.0055117	.2063598
	.005298	.0015354	0.13	0.011	.006411	.0059146
hsgrad	.1946563	.408713	0.48	0.634	-.606432	.9957447
	.0133429	.0021743	0.03	0.022	.0108107	.016606
female	-.1005675	.4181433	-0.24	0.810	-.9201222	.7189871
	.0104288	.0013308	0.03	0.020	.0094587	.0119029
_cons	-5.507377	1.73762	-3.17	0.002	-8.917259	-2.097494
	.1414801	.0332486	0.11	0.001	.1386845	.1728079

Note: Values displayed beneath estimates are Monte Carlo error estimates.

Steps of MI

1. Setup
2. Imputation
3. Analysis
4. Pooling
5. Postestimation

- Importing
- Data Management

Postestimation: Transformations

```
. mi estimate (diff:_b[smokes]-_b[bmi]), nocoef: ///
> logit attack smokes age bmi hsgrad female

Multiple-imputation estimates                   Imputations       =         20
Logistic regression                             Number of obs     =        154
                                                Average RVI       =     0.1381
                                                Largest FMI       =     0.1227
DF adjustment:   Large sample                   DF:     min       =   1,290.53
                                                        avg       =   1,290.53
Within VCE type:          OIM                           max       =   1,290.53

      command: logit attack smokes age bmi hsgrad female
         diff: _b[smokes]-_b[bmi]

------------------------------------------------------------------------------
      attack |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        diff |   1.125999   .3824437     2.94   0.003     .3757197    1.876279
------------------------------------------------------------------------------
```

Postestimation: Test

```
. mi test age hsgrad female
note: assuming equal fractions of missing information

 ( 1)  [attack]age = 0
 ( 2)  [attack]hsgrad = 0
 ( 3)  [attack]female = 0

       F(  3,186601.9) =    1.75
            Prob > F =    0.1546
```

Steps of MI

1. Setup
2. Imputation
3. Analysis
4. Pooling
5. Postestimation

- Importing
- Data Management

Importing

```
. import delimited heart_mi_unset, clear
(8 vars, 924 obs)

. mi import flong, m(imp) id(id) imputed(bmi) clear
(22 m=0 obs. now marked as incomplete)
```

Steps of MI

1. Setup
2. Imputation
3. Analysis
4. Pooling
5. Postestimation

- Importing
- Data Management

Data Management

- `mi append`
- `mi merge`
- `mi reshape`
- `mi extract` $\#$
- `mi xtset`
- `mi tsset`
- `mi svyset`
- `mi stset`
- `mi stsplit`
- `mi xeq:` *command*
- `mi passive: generate/egen/replace`
- For a full list type `help mi`

Generating Passive Variables

```
. mi passive: egen overweight = cut(bmi), at(0,25,40) icodes
m=0:
(22 missing values generated)
m=1:
m=2:
m=3:
m=4:
m=5:

. list bmi overweight _mi_m if id==4
```

	bmi	overwe~t	_mi_m
4.	.	.	0
158.	29.88588	1	1
312.	21.41766	0	2
466.	24.19195	0	3
620.	19.40182	0	4
774.	26.64958	1	5

Thank you!

Questions?

You can download the datasets and do-file here:
https://tinyurl.com/mi-web-2020

You can contact tech support at tech-support@stata.com