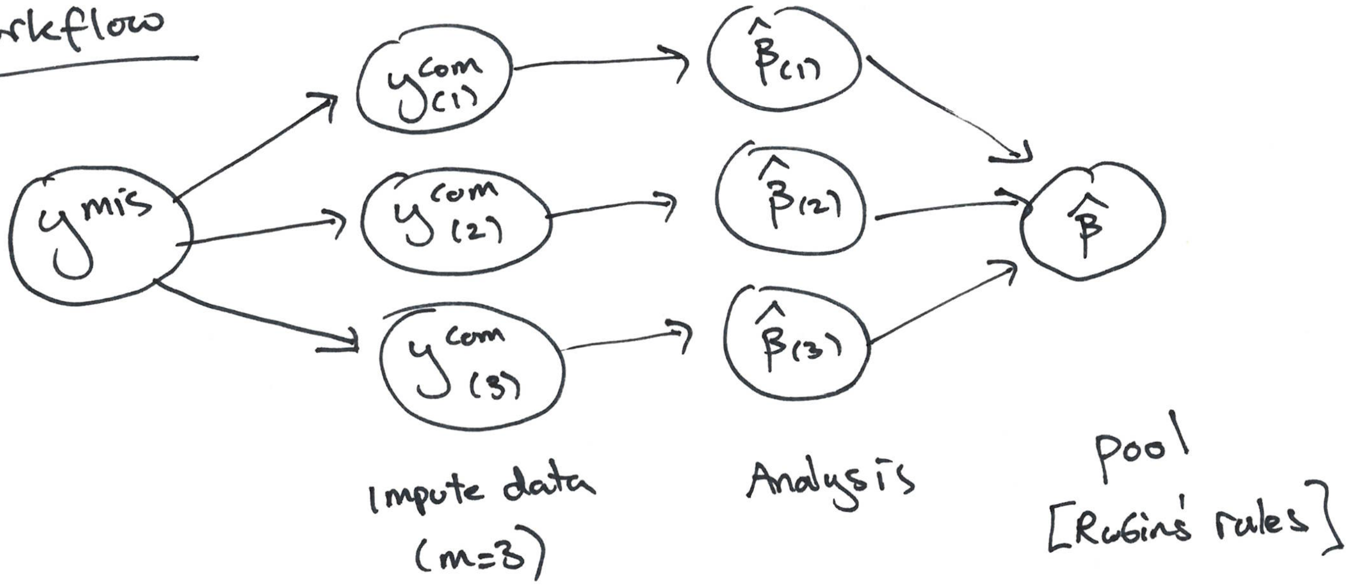
Multiple imputation by chained equations (MICE)

Specify an imputation model: $p(y^{mis} \mid y^{obs})$ where

$$y^{mis} = (y_1^{mis}, \ldots, y_p^{mis}), \quad y^{obs} = (y_1^{obs}, \ldots, y_p^{obs}).$$

Let $y^{com} = (y^{mis}, y^{obs})$ [combined / complete data with y^{mis} IMPUTED]

Workflow



To obtain a <u>single</u> value of the complete data, execute the following:

1. Initialise with $y^{mis(0)}$. Set $k=1$.
2. obtain $y^{mis(k)}$ from $y^{mis(k-1)}$ via
   - $y_1^{mis(k)} \sim p(\cdot \mid y^{obs}, y_2^{mis(k-1)}, \ldots, y_p^{mis(k-1)})$
   - $y_2^{mis(k)} \sim p(\cdot \mid y^{obs}, y_1^{mis(k)}, y_3^{mis(k-1)}, \ldots, y_p^{mis(k-1)})$

   ⋮
   - $y_p^{mis(k)} \sim p(\cdot \mid y^{obs}, y_1^{mis(k)}, \ldots, y_{p-1}^{(k)})$
3. $k \leftarrow k+1$, go to 2. Called a $\boxed{\text{Gibbs Sampler}}$

This workshop:

- Compare accuracy of imputation methods. How?
  - Take a data set, randomly delete values.
  - Let y_{mis} denote the true values of the missing data.
  - Let $\hat{y}_{mis}$ denote a single imputed value.

Compare accuracy of $\hat{y}_{mis}$ with reference to y_{mis} using:

- Mean absolute error (MAE)

$$MAE = \frac{1}{n}\sum_{i=1}^{n} \left| y_{mis,i} - \hat{y}_{mis,i} \right|$$

- Mean squared error (MSE)

$$MSE = \frac{1}{n}\sum_{i=1}^{n} \left(y_{mis,i} - \hat{y}_{mis,i} \right)^2$$

- Root mean squared error (RMSE)

$$RMSE = \sqrt{MSE}$$

- Mean absolute percentage error (MAPE)

$$MAPE = \frac{1}{n}\sum_{i=1}^{n} \left| \frac{y_{mis,i} - \hat{y}_{mis,i}}{y_{mis,i}} \right|$$
