Handling Ignorable and Non-ignorable Missing Data through Bayesian Methods in JAGS

Ziqian Xu

University of Notre Dame, Notre Dame, IN 46530, USA
zxu9@nd.edu

Abstract. With the prevalence of missing data in social science research, it is necessary to use methods for handling missing data. One framework in which data with missing value can still be used for parameter estimation is the Bayesian framework. In this tutorial, different missing data mechanisms including Missing Completely at Random, Missing at Random, and Missing Not at Random are introduced. Methods for estimating models with missing values under the Bayesian framework for both ignorable and non-ignorable missingness are also discussed. A structural equation model on data from the Advanced Cognitive Training for Independent and Vital Elderly study is used as an illustration on how to fit missing data models in JAGS.

Keywords: Missing data · Bayesian analysis · Structural equation model

1 Introduction

The problem of missing data is prevalent in research, and the social sciences are particularly influenced by this problem because surveys are commonly used to collect information. As pointed out by Berchtold (2019), item-level missing data were found in 69.5% of papers published in selected social science journals, suggesting that the issue of missing data is quite omnipresent. Missing data may lead to various problems including biases in estimations and lowered identifiability of models (e.g., Zhang & Wang 2012). To address missing data issues, procedures including listwise deletion, full-information maximum likelihood estimation, and multiple imputation have been proposed (e.g., Zhang & Wang 2013). One emerging context where models can be fitted with the presence of missing data is the Bayesian framework (Ma & Chen 2018). With Bayesian inference, missing data can be handled quite naturally. The purpose of this paper is to demonstrate the procedure of fitting a structural equation model with missing data under the Bayesian framework using the software JAGS (Plummer et al. 2003).
1.1 Missing Data Mechanisms

Three missing data mechanisms, Missing Completely at Random (MCAR), Missing at Random (MAR), and Missing Not at Random (MNAR), have been discussed in the literature (Rubin, 1976). To distinguish these mechanisms, we can use a vector $Y$ to denote the outcome variable, a matrix $X$ to denote the covariates, and a vector $R$ to denote the missing data indicator for the outcome $Y$, with $R = 1$ if a $Y$ is missing, otherwise 0. We can also assume a model with parameter $\gamma$ that governs the generation process of $R$. Further, we can assume that the purpose of data analysis is to obtain the parameter $\theta$ that explains $Y$ using $X$. To illustrate the different missing data mechanisms, we use missingness in the outcome $Y$ as an example.

- If the missing data in the outcome $Y$ are MCAR, the missing data do not depend on any data collected. Thus, $P(R|\gamma, Y, X) = P(R|\gamma)$.

- If the missing data in the outcome $Y$ are MAR, the missingness depends on the data collected, thus, $P(R|\gamma, Y, X) = P(R|\gamma, Y_{obs}, X_{obs})$ where $X_{obs}$ denotes covariates that are observed and used in the modeling processes to obtain $\theta$, and $Y_{obs}$ denotes the observed part of the outcome.

- If the missing data in the outcome $Y$ are MNAR, the missingness depends on the unobserved data, thus, $P(R|\gamma, Y, X) = P(R|\gamma, Y_{mis}, X_{mis}, Y_{obs}, X_{obs})$ where $X_{mis}$ denotes predictors or covariates that are not observed and $Y_{mis}$ denotes the missing outcome data.

1.2 Bayesian Methods for Handling Missing Outcome Data

**Ignorable Missing Data** Under the Bayesian framework, when the missing outcome data mechanism is ignorable (MCAR or MAR), meaning that the missingness can be viewed as random after accounting for observed data, the parameter $\theta$ of interest in the statistical analysis model can be estimated based on the observed part of data $Y_{obs}$. For simplicity, we assume covariates $X$ are fully observed for now. Thus, for ignorable missing outcome data, we can derive $P(\theta|Y_{obs}, X) \propto P(Y_{obs}|\theta, X)p(\theta)$ as the posterior of $\theta$ based on the observed part of data $P(Y_{obs}|\theta, X)$ and the prior $p(\theta)$. Any suitable model for the outcome variable can be used under the above scheme to obtain posterior samples of the parameter when missing data are present (Ma & Chen, 2018).

**Non-ignorable Missing Data** When the missing outcome data are non-ignorable (MNAR), meaning that the missingness mechanism cannot be fully explained by observed data, models such as the selection model and the pattern-mixture model can be used. Again, we assume that the covariates $X$ are fully observed for now. When missing data are non-ignorable, the target parameter of estimation is not only $\theta$, but both $\theta$ and $\gamma$. 
Selection Model. The selection model partitions the joint conditional probability of the outcome variable \(Y\) and missing data indicator \(R\) into two parts: 
\[
P(Y, R|\theta, \gamma, X) = P(Y|\theta, X) \cdot P(R|\gamma, Y, X) \quad \text{(Heckman 1979)}. 
\]
The part \(P(Y|\theta, X)\) denotes the probability of the outcome \(Y\) based on covariates \(X\) and parameters \(\theta\) not accounting for missingness. The part \(P(R|\gamma, Y, X)\) denotes the missingness mechanism based on both the covariates and the outcome data. While the selection model assumes the same analysis model for observed and missing data, it also assumes that the missingness indicator can be viewed as a function of the data. To incorporate the selection model into the Bayesian estimation process, we can write the posterior as 
\[
P(\theta, \gamma|Y, X, R) \propto P(Y|\theta, X) P(R|\gamma, X, Y) \pi(\theta, \gamma) 
\]
so that in addition to the model parameter \(\theta\), the missing data parameter \(\gamma\) is also estimated.

Pattern-Mixture Model. The pattern-mixture model factors the joint conditional probability of the outcome \(Y\) and the missingness indicator \(R\) in a different way: 
\[
P(Y, R|\theta, \gamma, X) = P(Y|\theta, X, R) \cdot P(R|\gamma, X) \quad \text{(Little 1994)}. 
\]
In this model, the part \(P(Y|\theta, X, R)\) indicates that the outcome \(Y\) depends on missingness (i.e., the outcome model could be different for observed and missing data resembling a mixture model), and the part \(P(R|\gamma, X)\) denotes the missingness mechanisms that only depend on the covariates and not on the outcome. To incorporate the pattern-mixture model into the Bayesian estimation process, we can express the posterior as 
\[
P(\theta, \gamma|Y, X, R) \propto P(Y|\theta, X, R) P(R|\gamma, X) \pi(\theta, \gamma). 
\]

1.3 Bayesian Methods for Handling of Missing Covariates Data

In Section 1.2, Bayesian schemes for handling missing data in the outcome variable are discussed assuming that the covariates are complete. However, covariates often contain missing data as well in real life. Thus, the estimation procedures need further adaptations. When the missing data mechanism for the covariates are ignorable (MCAR or MAR), distributions of the covariates can be specified in addition to the models in Section 1.2 such that each covariate can have its own conditional distribution [Ibrahim, Chen, & Lipsitz 2002]. When the missing covariates are non-ignorable (MNAR), models similar to the non-ignorable missing outcome model can also be applied to the covariates such as the selection model for both outcome and covariates implemented by [Lee and Tang 2006].

2 Data and Models

2.1 Data

A subset of data from the The Advanced Cognitive Training for Independent and Vital Elderly (ACTIVE) study will be used to illustrate how to handle missing data under the Bayesian framework. \(n = 500\) records for 5 variables, including 3 items measuring reasoning ability (i.e., \(WS, LS,\) and \(LT\)), \(AGE,\) and \(EPT\) will be used. Here, \(WS\) is the word series test result, \(LS\) is the letter series test result,
LT is the letter sets test result, AGE is the age of participants, and EPT is the Everyday Problems Test result. The respective scores are the integer number of correct answers in each test.

The subset of data selected contains no missing values. For the purpose of this tutorial, missing values are generated according to the MAR and MNAR mechanisms. For the ignorable missingness model, 14.8% of missing data in EPT are simulated such that \( \logit(P(R_i = 1)) = -0.6 + 0.01 \times \text{Age}_i \) where the \( R_i \) is the missingness indicator for the \( i \)-th record on EPT. For the selection model on handling non-ignorable missingness, 16.2% of missing data in EPT are simulated such that \( \logit(P(R_i = 1)) = 0.01 \times \text{EPT}_i \). Lastly, for the pattern mixture model on handing non-ignorable missingness, 17.6% of missing data in EPT are simulated by viewing EPT values between 24 and 26 as missing. Because EPT consists of integer values only, this means that only the EPT values of 24, 25 and 26 are missing. Thus, when viewing EPT as a linear function of other variables, which is what we will use in the analysis, the slopes for predictors will be small for missing EPT since there is very low variance in missing EPT compared to other variables; whereas for observed EPT, we can expect larger slopes. Additionally, because EPT values are mostly below 24, we can also expect the intercept for missing EPT to be larger than that of observed EPT.

2.2 Analysis Model

The analysis model will be a structural equation model using the three reasoning ability items (i.e., WS, LS, and LT) and AGE to predict EPT as shown in Figure 1. The measurement model can be written as in Equation 1.

\[
X_m = \mu + \eta \Lambda + \epsilon. \tag{1}
\]

We denote the three manifest variables that are the reasoning ability items involved in the measurement model using the matrix \( X_m \) with dimension \( n \times 3 \) as formatted in Equation 2. In addition, the latent variable \( \text{REASON} \) is represented by the second column in \( \eta_{n \times 1} \), the factor loadings are represented by \( \Lambda_{1 \times 3} \), the intercepts are represented by \( \mu_{1 \times 3} \), and the error terms are represented by \( \epsilon_{n \times 3} \). In the measurement model, it is assumed that \( \epsilon_{i1} \sim N(0, \sigma_{WS}^2) \), \( \epsilon_{i2} \sim N(0, \sigma_{LT}^2) \), and \( \epsilon_{i3} \sim N(0, \sigma_{LS}^2) \) for \( i \in (1, ..., n) \). Further, we assume the latent variable \( \text{REASON}_i \sim N(0, \sigma_{\text{REASON}}^2) \).
\[ X_m = \begin{bmatrix} WS_1 & LT_1 & LS_1 \\ \vdots & \vdots & \vdots \\ WS_n & LT_n & LS_n \end{bmatrix} \]
\[ \mu = \begin{bmatrix} k_1 \\ k_2 \\ k_3 \end{bmatrix} \]
\[ \eta = \begin{bmatrix} \text{REASON}_1 \\ \vdots \\ \text{REASON}_n \end{bmatrix} \]
\[ \Lambda = \begin{bmatrix} 1 & \lambda_{LT} & \lambda_{LS} \end{bmatrix} \]
\[ \epsilon = \begin{bmatrix} \epsilon_{11} & \epsilon_{12} & \epsilon_{13} \\ \vdots & \vdots & \vdots \\ \epsilon_{n1} & \epsilon_{n2} & \epsilon_{n3} \end{bmatrix} \]

**Figure 1.** Analysis model for the ACTIVE data.

The structural model follows Equation 3. Since the outcome is one-dimensional, we write the structural model with respect to the outcome \( EPT \) directly. In the structural model, the latent variable \( \text{REASON} \) and the observed variable \( \text{AGE} \) are used to predict \( EPT \) with an intercept. Similar to the assumptions of the measurement model, it is assumed that \( \epsilon_{EPT} \sim N(0, \sigma_{EPT}^2) \).

\[ EPT_i = b_0 + b_1 \text{REASON}_i + b_2 \text{AGE}_i + \epsilon_{EPT}^i. \]  

(3)

Together, the measurement and structural models can also be written with respect to each item as in Equation 4.
\[ WS_i = k_{WS} + REASON_i + \epsilon_i \]
\[ LT_i = k_{LT} + \lambda_{LT} REASON_i + \epsilon_i \]
\[ LS_i = k_{LS} + \lambda_{LS} REASON_i + \epsilon_i \] (4)
\[ EPT_i = b_0 + b_1 REASON_i + b_2 AGE_i + \epsilon_{EPT} \]

Combining the measurement model and the structural model, the probability distribution function form of the analysis model used can be written as in Equation 5:

\[ REASON_i \sim N(0, \sigma_{REASON}^2) \]
\[ WS_i \sim N(k_{WS} + REASON_i, \sigma_{WS}^2) \]
\[ LT_i \sim N(k_{LT} + \lambda_{LT} REASON_i, \sigma_{LT}^2) \]
\[ LS_i \sim N(k_{LS} + \lambda_{LS} REASON_i, \sigma_{LS}^2) \]
\[ EPT_i \sim N(b_0 + b_1 REASON_i + b_2 AGE_i, \sigma_{EPT}^2) \] (5)

2.3 Software

JAGS will be used to conduct Bayesian inference in this paper. JAGS can be installed via https://mcmc-jags.sourceforge.io/ (Plummer et al., 2003). In this paper, JAGS will be used in the R environment (R Core Team, 2021) through the software RStudio (RStudio Team, 2022) and the package runjags (Denwood, 2016). Other packages such as rjags (Plummer, Stukalov, & Denwood, 2022) that provide similar functionalities to runjags are available as well. JAGS can also be executed from the command line without using another interface.

In R, the package runjags can be installed using `install.packages("runjags")` and loaded using `library(runjags)`.

3 Ignorable Missingness Model

3.1 Model Specification

In Bayesian inference with JAGS, the two crucial components to a model are the likelihood and the priors.

**Likelihood** If the missingness in the outcome of \( EPT \) is ignorable in the data, then missing data can be addressed by sampling from the analysis model using observed data. Thus, the model in Equation 4 can be used as the ignorable missingness model. Further, the likelihood can be specified using the probability densities in Equation 5. In this example, we do not have missing data in the covariates. However, when missing data are present in the covariates, additional distributions will need to be specified for them.
Priors We also need to specify the priors for the ignorable missingness model. Since we are using the assumption that the observed and latent variables all follow normal distributions, which is common in the practice of structural equation modeling, we can choose the priors in Equation 6 for the model parameters. In particular, the regression coefficients, factor loadings, and intercepts will have normal priors, and the variance components will have inverse gamma priors. Because there is only one latent factor in the model used, the inverse gamma distribution could satisfy as the prior for the factor variance. However, when more than one factors are involved in a structural equation model, and when the factors are allowed to correlate, the inverse wishart distribution can be used as the prior for the factor covariance matrix.

\[
\begin{align*}
b_0, b_1, b_2 &\sim N(\mu_b, \sigma^2_b) \\
\sigma^2_{\text{EPT}} &\sim IG(h_{\text{EPT}}, \theta_{\text{EPT}}) \\
\lambda_{LT}, \lambda_{LS} &\sim N(\mu_\lambda, \sigma^2_\lambda) \\
\sigma^2_{\text{WS}, \text{LT}}, \sigma^2_{\text{WS}, \text{LS}} &\sim IG(h_m, \theta_m) \\
k_{WS}, k_{LT}, k_{LS} &\sim N(\mu_k, \sigma^2_k) \\
\sigma^2_{\text{REASON}} &\sim IG(h_{\text{REASON}}, \theta_{\text{REASON}})
\end{align*}
\]

In this tutorial, we will choose rather uninformative priors. But more informative priors can also be adopted. For example, the power priors can be used which constructs priors based on likelihood in historical data (Ibrahim & Chen, 2000). Hierarchical priors is another option which can be used when a range of values for the priors are available, so different levels of priors can be specified (Berger & Strawderman, 1996). These methods can be utilized for more theory-informed specification of priors.

3.2 Implementation in JAGS

Implementing a model in JAGS involves the following steps.

Model Specification in JAGS Using runjags, the model specification can be stored as a string. For example, the code below specifies a model named ignorable.

```r
ignorable <- "
#likelihood
... #prior
...
"
```

The likelihood as specified in Section 3.1 can be implemented inside the model string as the following. In JAGS, distributions can be specified using the symbol ~. As an illustration, in the code EPT[i] ~ dnorm(mu.EPT[i], pre.EPT),
dnorm indicates that \( EPT \) is expected to follow a normal distribution with the mean of \( \mu.EPT \) and the precision (i.e., inverse of variance) of \( \text{pre.EPT} \). Further, \( \mu.EPT[i] \leftarrow b0 + b1*\text{REASON}[i] + b2*\text{AGE}[i] \) shows that \( EPT \) is predicted by the latent variable \( \text{REASON} \) and the observed variable \( \text{AGE} \). The two lines of code together correspond to Equation 3. Note that while ~ is used to specify a distribution, <- is used to assign values. Similar to how the distribution of \( EPT \) is specified, the measurement model for the 3 reasoning ability items can be specified.

```r
# likelihood
for (i in 1:N) {
  EPT[i] ~ dnorm(mu.EPT[i], pre.EPT)
  mu.EPT[i] <- b0 + b1*REASON[i] + b2*AGE[i]
  WS[i] ~ dnorm(mu1[i], pre.WS)
  LT[i] ~ dnorm(mu2[i], pre.LT)
  LS[i] ~ dnorm(mu3[i], pre.LS)
  mu1[i] <- REASON[i] + k.WS
  mu2[i] <- l.LT*REASON[i] + k.LT
  mu3[i] <- l.LS*REASON[i] + k.LS
  REASON[i] ~ dnorm(0, pre.REASON)
}
```

As mentioned in Section 3.1, historical data and previous research conclusions are not used to construct the priors here; instead, the priors used are rather uninformative. For example, the prior for means the error terms in the factor model of \( \text{REASON} \) has precision of 0.001 or variance of 1000.

```r
# priors
# regression model
b0 ~ dnorm(0, pre.b)
b1 ~ dnorm(0, pre.b)
b2 ~ dnorm(0, pre.b)
pre.b ~ dgamma(.001, .001)
pre.EPT ~ dgamma(0.001, 0.001)

# factor model
l.LT ~ dnorm(0, 0.001)
l.LS ~ dnorm(0, 0.001)
pre.WS ~ dgamma(0.001, 0.001)
pre.LT ~ dgamma(0.001, 0.001)
pre.LS ~ dgamma(0.001, 0.001)
k.WS ~ dnorm(0, 0.001)
k.LT ~ dnorm(0, 0.001)
k.LS ~ dnorm(0, 0.001)
pre.REASON ~ dgamma(0.001, 0.001)
```

Additionally, to monitor the variances of various variables instead of the precision, we can use the following to convert precisions to variances.
# variances
var.EPT <- 1/pre.EPT
var.WS <- 1/pre.WS
var.LT <- 1/pre.LT
var.LS <- 1/pre.LS
var.REASON <- 1/pre.REASON

Data Statement and Initial Values To estimate the ignorable missingness model, we define the data to be used as follows.

data <- list(N = 500,
EPT = data_500_mar$EPT,
AGE = data_500_mar$AGE,
WS = data_500_mar$WS,
LT = data_500_mar$LT,
LS = data_500_mar$LS)

We also define initial values for the two MCMC chains that will be used. Slightly different starting values for the two chains are specified to ensure that the two chains will not converge to the same local optimum, if happening, for better convergence diagnostic.

inits <- list(list(b0=0, b1=0, b2=0, l.LT=0, l.LS=0,
k.WS=0, k.LS=0, k.LT=0,
pre.WS=0.1, pre.LS=0.1, pre.LT=0.1,
pre.EPT=0.1, pre.REASON=0.1),
list(b0=1, b1=1, b2=1, l.LT=1, l.LS=1,
k.WS=1, k.LS=1, k.LT=1,
pre.WS=0.2, pre.LS=0.2, pre.LT=0.2,
pre.EPT=0.2, pre.REASON=0.2))

Running JAGS for Analysis Finally, the posteriors can be sampled using the run.jags command. The model=ignorable argument specifies that the model named m1 will be used. The monitor argument specifies which parameters or variables we want to sample. The data=data, n.chains=2, inits=inits arguments specify the data, number of chains, and initial values to be used, respectively. The arguments adapt=1000, burnin = 3000, and sample=100000 specify the number of adaptation, burn-in, and sampling iterations, respectively. Only the sampling iterations will be recorded. The argument keep.jags.files=FALSE instructs JAGS to not save any files produced in the sampling process. If the argument is set to keep.jags.files=FALSE, then additional files will be saved in a folder named runjagsfiles. The argument thin=1 sets the intervals of recorded samples to be 1, meaning that every iteration will be saved. The argument method="simple" indicates that the simple method would be used to compile
the model. The argument `tempdir=TRUE` asks JAGS to create a temporary directory to store files instead of saving files in the working directory.

```r
out1 <- run.jags(model=ignorable,
    monitor=c("b0", "b1", "b2",
              "l.LT", "l.LS",
              "k.WS", "k.LT", "k.LS",
              "var.WS", "var.LT",
              "var.LS", "var.EPT", "var.REASON"),
    data=data, n.chains=2,
    inits=inits, method="simple",
    adapt=1000, burnin = 3000,
    sample=1000000,
    thin=1,
    keep.jags.files=FALSE,
    tempdir=TRUE)
```

### 3.3 Convergence Diagnostics

The `summary(out1)` command in `runjags` gives the summary statistics of the JAGS output which can be used for convergence diagnostics. We will mainly rely on the Gelman-Rubin test for diagnosing convergence (Gelman & Rubin, 1992). The Gelman-Rubin test statistics, or the potential scale reduction factor, is readily provided by the `summary` function and is suitable when multiple chains are used. As shown in Table 1, the potential scale reduction factors (psrf) for all parameters are below 1.1 and very close to 1, suggesting that the chains have converged. Monte Carlo SEs of most parameters (MCerr) are quite low as well (< 0.05). The effective sample sizes (SSeff) are also acceptable (> 400).

#### Table 1. Output from the ignorable data model.

<table>
<thead>
<tr>
<th></th>
<th>Lower95</th>
<th>Median</th>
<th>Upper95</th>
<th>Mean</th>
<th>MCerr</th>
<th>SSeff</th>
<th>perf</th>
</tr>
</thead>
<tbody>
<tr>
<td>b0</td>
<td>10.546</td>
<td>16.071</td>
<td>21.450</td>
<td>16.045</td>
<td>0.041</td>
<td>4515</td>
<td>1</td>
</tr>
<tr>
<td>b1</td>
<td>0.855</td>
<td>0.960</td>
<td>1.062</td>
<td>0.960</td>
<td>0.000</td>
<td>16217</td>
<td>1</td>
</tr>
<tr>
<td>b2</td>
<td>-0.038</td>
<td>0.034</td>
<td>0.110</td>
<td>0.034</td>
<td>0.001</td>
<td>4485</td>
<td>1</td>
</tr>
<tr>
<td>l.LT</td>
<td>0.402</td>
<td>0.445</td>
<td>0.489</td>
<td>0.445</td>
<td>0.000</td>
<td>20000</td>
<td>1</td>
</tr>
<tr>
<td>l.LS</td>
<td>1.066</td>
<td>1.142</td>
<td>1.222</td>
<td>1.142</td>
<td>0.000</td>
<td>20000</td>
<td>1</td>
</tr>
<tr>
<td>k.WS</td>
<td>9.266</td>
<td>9.700</td>
<td>10.129</td>
<td>9.700</td>
<td>0.002</td>
<td>20000</td>
<td>1</td>
</tr>
<tr>
<td>k.LT</td>
<td>5.439</td>
<td>5.684</td>
<td>5.926</td>
<td>5.683</td>
<td>0.001</td>
<td>20000</td>
<td>1</td>
</tr>
<tr>
<td>k.LS</td>
<td>9.744</td>
<td>10.225</td>
<td>10.726</td>
<td>10.226</td>
<td>0.002</td>
<td>20483</td>
<td>1</td>
</tr>
<tr>
<td>var.WS</td>
<td>3.294</td>
<td>4.247</td>
<td>5.265</td>
<td>4.260</td>
<td>0.004</td>
<td>20000</td>
<td>1</td>
</tr>
<tr>
<td>var.LT</td>
<td>3.025</td>
<td>3.487</td>
<td>3.995</td>
<td>3.497</td>
<td>0.002</td>
<td>20000</td>
<td>1</td>
</tr>
<tr>
<td>var.LS</td>
<td>3.717</td>
<td>4.981</td>
<td>6.230</td>
<td>4.994</td>
<td>0.005</td>
<td>19548</td>
<td>1</td>
</tr>
<tr>
<td>var.EPT</td>
<td>12.649</td>
<td>14.831</td>
<td>17.112</td>
<td>14.884</td>
<td>0.008</td>
<td>20000</td>
<td>1</td>
</tr>
<tr>
<td>var.REASON</td>
<td>17.037</td>
<td>19.849</td>
<td>23.122</td>
<td>19.903</td>
<td>0.011</td>
<td>20000</td>
<td>1</td>
</tr>
</tbody>
</table>
The `plot(out1)` command can be used to generate trace plots and histograms of the MCMC chains which can further help diagnose the convergence of the target parameters. For example, Figure 2 shows the MCMC plots of the parameter $b_1$, and the trace plot and histogram both indicate that the parameter converged. The histogram is centered at one mode, and the trace plot looks stable.

**Figure 2.** MCMC plots for the parameter $b_1$ in the ignorable missingness model.

There are other approaches for diagnosing convergence which will not be discussed in depth here. For example, the Geweke test can be conducted using the `coda` package (Plummer, Best, Cowles, & Vines, 2006). This test takes two proportions of a Markov chain and applies a z-test to see if the two proportions are different. The Herdelberger and Welch test can also be used with `coda` which tests if the samples come from a stable distribution.

### 3.4 Interpretation

Using the Highest Posterior Density (HPD) intervals which are indicated by the Lower95 and Upper95 values in Table 1, we can see that the factor loadings and intercepts in the measurement model are likely non-zero as the HPD intervals for them excludes 0. Thus, the items $WS$, $LT$ and $LS$ do have commonalities explained by the construct of reasoning ability. In the structural model, the intercept term and the slope for $REASON$ have their HPD intervals excluding 0, suggesting that we can conclude that the latent variable of $REASON$ explains the variance in $EPT$, whereas $AGE$ does not.
4 Selection Model

4.1 Model Specification

**Likelihood** When the missingness in the outcome is MNAR or non-ignorable, the selection model can be applied. For non-ignorable missing outcome, we also make the same assumption that the missing covariates are ignorable so that we can specify conditional distributions for them. We also assume that the missingness is dependent on the outcome of EPT scores themselves in the selection model. Using a missing data indicator $R_i$ where $R_i = 1$ denotes a missing record of EPT and $R_i = 0$ denotes an observed record of EPT, the selection model translates to Equation 7 which can be used in addition to the model specification in Equation 4:

$$\logit(P(R_i = 1)) = a_0 + a_1 \text{EPT}_i.$$  

(7)

The corresponding probability density distribution form of the selection model incorporates Equation 8 in addition to Equation 5:

$$R_i \sim \text{Bernoulli}(\text{sigmoid}(a_0 + a_1 \text{EPT}_i)).$$

(8)

The missing data indicator $R_i$ is deemed as a function of the outcome value $\text{EPT}_i$ and an intercept. The intercept $a_0$ here is used to adjust the threshold for 1 and 0 in the missing data indicator of $R_i$.

**Priors** The priors can be constructed similarly to the ignorable missingness model. Priors for the parameters $a_0$ and $a_1$ are needed for the selection model as indicated by Equation 9 which can be combined with the priors in Equation 6 to be used in JAGS.

$$a_0, a_1 \sim \mathcal{N}(\mu_a, \sigma_a^2)$$

(9)

4.2 Implementation in JAGS

**Model Specification in JAGS** In JAGS, the likelihood for the ignorable data model can be modified to incorporate the selection model by including the following.

$$R[i] \sim \text{dbern}(p[i])$$
$$\text{logit}(p[i]) = a_0 + a_1 \text{EPT}[i]$$

The priors also need to be modified based on the the function of missingness on $\text{EPT}$.

$$a_0 \sim \text{dnorm}(0, \text{pre.a})$$
$$a_1 \sim \text{dnorm}(0, \text{pre.a})$$
$$\text{pre.a} \sim \text{dgamma}(0.001, 0.001)$$
Data Statement and Initial Values Since the model has now changed, the initial values and data supplied should be modified accordingly. A missing data indicator for the outcome is also needed as shown below in the variable $R = \text{is.na(data\_500\_mnar\$EPT)} \times 1$.

```r
data <- list(N = 500,
EPT = data\_500\_mnar\$EPT,
AGE = data\_500\_mnar\$AGE,
WS = data\_500\_mnar\$WS,
LT = data\_500\_mnar\$LT,
LS = data\_500\_mnar\$LS,
R = \text{is.na(data\_500\_mnar\$EPT)} \times 1)
```

```r
inits <- list(list(b0=0, b1=0, b2=0, l.LT=0, l.LS=0,
k.WS=0, k.LS=0, k.LT=0,
pre.WS=0.1, pre.LS=0.1, pre.LT=0.1,
pre.EPT=0.1, pre.REASON=0.1,
a0=0.1, a1=0.1),
list(b0=1, b1=1, b2=1, l.LT=1, l.LS=1,
k.WS=1, k.LS=1, k.LT=1,
pre.WS=0.2, pre.LS=0.2, pre.LT=0.2,
pre.EPT=0.2, pre.REASON=0.2,
a0=0.2, a1=0.2))
```

Running JAGS for Analysis The model can then be estimated using the following statement. Again, 1000 adaptation iterations, 3000 burn-in iterations, and 100000 sampling iterations are specified.

```r
out1 <- run.jags(model=selection,
monitor=c("b0", "b1", "b2",
"a0", "a1",
"l.LT", "l.LS", "k.WS",
"k.LT", "k.LS",
"var.WS", "var.LT",
"var.LS", "var.EPT", "var.REASON"),
data=data, n.chains=2,
inits=inits, method="simple",
adapt=1000, burnin = 3000,
sample=1000000,
thin=1,
keep.jags.files=FALSE,
tempdir=TRUE)
```
4.3 Convergence Diagnostics

As shown in Table 2, the potential scale reduction factors (psrf) are all below 1.1, and the Monte Carlo SEs (MCerr) are satisfactory (< 0.05). The effective sample sizes (SSeff) for all parameters are also above 400. The trace plots are satisfactory as well. As an example, MCMC plots for the parameter $a_1$ is shown in Figure 3.

Table 2. Output from the selection model.

<table>
<thead>
<tr>
<th></th>
<th>Lower95</th>
<th>Median</th>
<th>Upper95</th>
<th>Mean</th>
<th>MCerr</th>
<th>SSeff</th>
<th>psrf</th>
</tr>
</thead>
<tbody>
<tr>
<td>$b_0$</td>
<td>13.170</td>
<td>18.649</td>
<td>24.029</td>
<td>18.623</td>
<td>0.043</td>
<td>4048</td>
<td>1</td>
</tr>
<tr>
<td>$b_1$</td>
<td>0.865</td>
<td>0.968</td>
<td>1.077</td>
<td>0.968</td>
<td>0.000</td>
<td>15546</td>
<td>1</td>
</tr>
<tr>
<td>$b_2$</td>
<td>-0.074</td>
<td>-0.002</td>
<td>0.072</td>
<td>-0.002</td>
<td>0.001</td>
<td>4064</td>
<td>1</td>
</tr>
<tr>
<td>$a_0$</td>
<td>-4.276</td>
<td>-2.752</td>
<td>-1.374</td>
<td>-2.788</td>
<td>0.006</td>
<td>13872</td>
<td>1</td>
</tr>
<tr>
<td>$a_1$</td>
<td>-0.010</td>
<td>0.059</td>
<td>0.132</td>
<td>-0.002</td>
<td>0.001</td>
<td>13990</td>
<td>1</td>
</tr>
<tr>
<td>$l_{LT}$</td>
<td>0.402</td>
<td>0.446</td>
<td>0.491</td>
<td>0.446</td>
<td>0.000</td>
<td>20000</td>
<td>1</td>
</tr>
<tr>
<td>$l_{LS}$</td>
<td>1.076</td>
<td>1.153</td>
<td>1.233</td>
<td>1.154</td>
<td>0.000</td>
<td>20551</td>
<td>1</td>
</tr>
<tr>
<td>$k_{WS}$</td>
<td>9.266</td>
<td>9.702</td>
<td>10.133</td>
<td>9.702</td>
<td>0.002</td>
<td>20000</td>
<td>1</td>
</tr>
<tr>
<td>$k_{LT}$</td>
<td>5.445</td>
<td>5.685</td>
<td>5.926</td>
<td>5.685</td>
<td>0.001</td>
<td>19607</td>
<td>1</td>
</tr>
<tr>
<td>$k_{LS}$</td>
<td>9.747</td>
<td>10.230</td>
<td>10.720</td>
<td>10.228</td>
<td>0.002</td>
<td>19719</td>
<td>1</td>
</tr>
<tr>
<td>$var_{WS}$</td>
<td>3.436</td>
<td>4.448</td>
<td>5.453</td>
<td>4.459</td>
<td>0.004</td>
<td>20000</td>
<td>1</td>
</tr>
<tr>
<td>$var_{LT}$</td>
<td>3.048</td>
<td>3.502</td>
<td>4.014</td>
<td>3.513</td>
<td>0.002</td>
<td>21125</td>
<td>1</td>
</tr>
<tr>
<td>$var_{LS}$</td>
<td>3.443</td>
<td>4.699</td>
<td>5.955</td>
<td>4.711</td>
<td>0.005</td>
<td>20000</td>
<td>1</td>
</tr>
<tr>
<td>$var_{EPT}$</td>
<td>12.945</td>
<td>15.249</td>
<td>17.670</td>
<td>15.308</td>
<td>0.009</td>
<td>19667</td>
<td>1</td>
</tr>
<tr>
<td>$var_{REASON}$</td>
<td>16.618</td>
<td>19.644</td>
<td>22.728</td>
<td>19.701</td>
<td>0.011</td>
<td>20000</td>
<td>1</td>
</tr>
</tbody>
</table>

4.4 Interpretation

Using the HPD intervals, the factor loadings and factor intercepts in the measurement model are again non-zero, indicating that the three items measuring \textit{REASON} do have overlaps. In the structural part of the selection model, the intercept and the slope for the latent variable \textit{REASON} both have HPD intervals excluding 0, suggesting that \textit{REASON} do explain significant variance in \textit{EPT} after accounting for the missingness mechanism. However, the HPD interval for the slope of \textit{AGE} still contains 0, suggesting that \textit{AGE} may not be a useful predictor of \textit{EPT} here.

The slope for \textit{EPT} in explaining the missingness mechanism has HPD intervals containing 0, suggesting that \textit{EPT}’s effect on explaining missingness is small. Whereas the intercept for predicting missingness has its HPD interval excluding 0. This deviates from the data generation model because the coefficients used in missing data generation are quite small.

Although the selection model is most commonly used when missing data are non-ignorable, it can be applied to MCAR and MAR data as well. In the ACTIVE example, missingness predictors can be changed to, for instance,
logit(p[i]) = a0 for MCAR missingness and logit(p[i]) = a0 + a1*AGE[i] for MAR missingness, for example.

Figure 3. MCMC plots for the parameter a1 in the selection model.

5 Pattern-Mixture Model

5.1 Model Specification

**Likelihood** A pattern-mixture model can also be used to handle non-ignorable missing data in the outcome of \( EPT \). In this case, the intercept \( b_0 \) and the slope \( b_1 \) for the outcome regressed on the latent variable of \( REASON \) are deemed different for missing and observed data, whereas in practice, theories should be used to guide the decision on the missing data model. The pattern-mixture property is shown as the subscripts \( R_i \) on \( b_0 \) and \( b_1 \) in Equation 10, which can be used in addition to the model specification in Equation 4 as the specification for the pattern-mixture model:

\[
EPT_i = b_{0R_i} + b_{1R_i} \cdot REASON_i + b_2AGE_i + \epsilon_i^{EPT}. \tag{10}
\]

Also, Equation 11 can be used in addition to Equation 5 as the probability density function form of the pattern-mixture model:

\[
EPT_i \sim N(b_{0R_i} + b_{1R_i} \cdot REASON_i + b_2AGE_i, \sigma_y^2). \tag{11}
\]

Assuming \( R_i = 1 \) for a missing record on \( EPT \) and \( R_i = 0 \) for an observed record on \( EPT \), then there are two sets of distinct \( b_0 \) and \( b_1 \) values for \( R_i = 1 \) and
$R_i = 0$, respectively. It should be noted that this is merely one possible pattern-mixture model that can be fitted onto the data, and that other pattern-mixture models can be used based on specific theories.

**Priors** Priors for the pattern-mixture model can be specified using Equation [12] to replace the corresponding regression coefficients priors in Equation [6]. Different from the priors for the ignorable missingness model, here, priors for $b_0$ and $b_1$ are differentiated for missing data ($b_0^1, b_1^1$) and for observed data ($b_0^0, b_1^0$).

$$b_0^0, b_1^0, b_0^1, b_1^1 \sim N(\mu_b, \sigma_b^2) \quad (12)$$

### 5.2 Implementation in JAGS

**Model Specification in JAGS** Different from the ignorable missingness model, we need to specify the mean of $EPT$ to be different for missing and observed data such as $mu.EPT[i] \leftarrow b_0[R[i]+1] + b_1[R[i]+1] \times REASON[i] + b_2[AGE][i]$ in JAGS. Since JAGS uses 1-based indexes instead of 0-based indexes, we need to use $R[i]+1$ instead of $R[i]$ in the code to transform the 0 and 1 values in the missing value indicator to 1 and 2. We also need to modify the priors of $b_0$ and $b_1$ as the following which is based on the assumption that while the intercept is larger for missing $EPT$, the slope is smaller for missing $EPT$. This assumption is based on how missing data is generated.

- $b0[1] \sim dnorm(0, pre.b)$ # non-missing
- $b0[2] \sim dnorm(b0[1]+0.5, pre.b)$ # missing
- $b1[1] \sim dnorm(0, pre.b)$ # non-missing
- $b1[2] \sim dnorm(b1[1]-0.5, pre.b)$ # missing

**Data Statement and Initial Values** Similar to the case in the selection model, an indicator for missing or observed data is required for estimating the pattern-mixture model, and the data statement can take the form of the following:

```r
data <- list(N = 500,
EPT = data_500_mnar_2$EPT,
AGE = data_500_mnar_2$AGE,
WS = data_500_mnar_2$WS,
LT = data_500_mnar_2$LT,
LS = data_500_mnar_2$LS,
R = is.na(data_500_mnar_2$EPT)*1)
```

The initial values can be specified based on the pattern-mixture model. Compared to the ignorable missingness model, initial values for $b_0$ and $b_1$ are changed to incorporate two cases for missing and observed data. For example, in the first chain, initial values for $b_0$ and $b_1$ are $b0=c(0,1), b1=c(1,0)$. This is consistent with our expectation that the intercept and slope for missing data would be greater and lower, respectively, than those of the observed data.
JAGS missing data

```r
inits <- list(list(b0=c(0,1), b1=c(1,0), b2=0,
    l.LT=0, l.LS=0, k.WS=0, k.LS=0, k.LT=0,
    pre.WS=0.1, pre.LS=0.1, pre.LT=0.1, pre.EPT=0.1, pre.REASON=0.1),
    list(b0=c(1,2), b1=c(2,1), b2=1,
    l.LT=1, l.LS=1, k.WS=1, k.LS=1, k.LT=1,
    pre.WS=0.2, pre.LS=0.2, pre.LT=0.2, pre.EPT=0.2, pre.REASON=0.2))
```

Running JAGS for Analysis Finally, the model can be estimated using the statement below.

```r
out1 <- run.jags(model=pmm,
    monitor=c("b0", "b1", "b2",
    "l.LT", "l.LS",
    "k.WS", "k.LT", "k.LS", "var.WS", "var.LT",
    "var.LS", "var.EPT", "var.REASON"),
    data=data, n.chains=2,
    inits=inits, method="simple",
    adapt=1000, burnin = 3000,
    sample=1000000,
    thin=1,
    keep.jags.files=FALSE,
    tempdir=TRUE)
```

5.3 Convergence Diagnostics

In Table 3 which contains the output from this model, the potential scale reduction factors (psrf) are lower than 1.1, suggesting that the chains have converged. The effective sample sizes (SSeff) are acceptable (> 400), and the Monte Carlo SEs (MCerr) are mostly below 0.05.

The MCMC plots in Figure 4 can also be used to diagnose convergence. Here, the plot for $b_1$ when data are observed is used as an example. All parameters, similar to $b_1$, present satisfactory trace plots suggesting that convergence is reached and the estimates are stable.

5.4 Interpretation

As shown by the HPD intervals in Table 3, the factor loadings and intercepts in the measurement model again exhibit HPD intervals that exclude 0. In the structural model, $AGE$’s HPD interval contains 0, whereas the intercept and the slope for $REASON$ have HPD intervals excluding 0 for observed data and including 0 for missing data. Consequently, for observed data, $REASON$ can predict $EPT$ whereas for missing data it is not the case.
Table 3. Output from the pattern-mixture model.

<table>
<thead>
<tr>
<th></th>
<th>Lower95</th>
<th>Median</th>
<th>Upper95</th>
<th>Mean</th>
<th>MCerr</th>
<th>SSeff</th>
<th>psrf</th>
</tr>
</thead>
<tbody>
<tr>
<td>$b_0[2]$</td>
<td>-7.595</td>
<td>21.045</td>
<td>52.908</td>
<td>21.498</td>
<td>0.113</td>
<td>20000</td>
<td>1.00</td>
</tr>
<tr>
<td>$b_1[1]$</td>
<td>0.847</td>
<td>0.950</td>
<td>1.053</td>
<td>0.951</td>
<td>0.000</td>
<td>17588</td>
<td>1.00</td>
</tr>
<tr>
<td>$b_1[2]$</td>
<td>-30.761</td>
<td>0.527</td>
<td>29.239</td>
<td>0.586</td>
<td>0.119</td>
<td>20000</td>
<td>1.01</td>
</tr>
<tr>
<td>$b_2$</td>
<td>-0.046</td>
<td>0.020</td>
<td>0.084</td>
<td>0.020</td>
<td>0.000</td>
<td>5211</td>
<td>1.00</td>
</tr>
<tr>
<td>$l.LT$</td>
<td>0.399</td>
<td>0.443</td>
<td>0.486</td>
<td>0.443</td>
<td>0.000</td>
<td>20000</td>
<td>1.00</td>
</tr>
<tr>
<td>$l.LS$</td>
<td>1.066</td>
<td>1.141</td>
<td>1.222</td>
<td>1.142</td>
<td>0.000</td>
<td>19925</td>
<td>1.00</td>
</tr>
<tr>
<td>k.WS</td>
<td>9.288</td>
<td>9.705</td>
<td>10.149</td>
<td>9.705</td>
<td>0.002</td>
<td>20328</td>
<td>1.00</td>
</tr>
<tr>
<td>k.LT</td>
<td>5.450</td>
<td>5.686</td>
<td>5.925</td>
<td>5.686</td>
<td>0.001</td>
<td>19697</td>
<td>1.00</td>
</tr>
<tr>
<td>k.LS</td>
<td>9.741</td>
<td>10.232</td>
<td>10.716</td>
<td>10.231</td>
<td>0.002</td>
<td>20291</td>
<td>1.00</td>
</tr>
<tr>
<td>var.WS</td>
<td>3.254</td>
<td>4.204</td>
<td>5.240</td>
<td>4.218</td>
<td>0.004</td>
<td>20000</td>
<td>1.00</td>
</tr>
<tr>
<td>var.LT</td>
<td>3.070</td>
<td>3.524</td>
<td>4.036</td>
<td>3.533</td>
<td>0.002</td>
<td>20000</td>
<td>1.00</td>
</tr>
<tr>
<td>var.LS</td>
<td>3.711</td>
<td>4.955</td>
<td>6.239</td>
<td>4.968</td>
<td>0.005</td>
<td>20000</td>
<td>1.00</td>
</tr>
<tr>
<td>var.EPT</td>
<td>11.269</td>
<td>13.222</td>
<td>15.379</td>
<td>13.268</td>
<td>0.007</td>
<td>20000</td>
<td>1.00</td>
</tr>
<tr>
<td>var.REASON</td>
<td>16.956</td>
<td>19.852</td>
<td>23.114</td>
<td>19.932</td>
<td>0.011</td>
<td>20000</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Figure 4. MCMC plots for the parameter $b_1$ for observed $EPT$ data in the pattern-mixture model.
6 Discussion

Data with missing values can still be used for parameter estimation in different models under the Bayesian framework. In this paper, a structural equation model is used as an example on the ACTIVE study data to illustrate how models with missing data can be fitted using JAGS in R under different missing data mechanisms including MCAR, MAR, and MNAR. Specifically, two models, the selection model and the pattern-mixture model, are introduced for non-ignorable (i.e., MNAR) missing data.

While this paper mainly focuses on obtaining parameter estimates when missing data are present, other aspects of missing data are not covered here. Although not discussed in this paper, other models for handling non-ignorable missingness exist such as the shared-parameter model (Albert & Follmann 2008). In addition, there are model selection methods for deciding which of many potential missing data models would suit the data best. For example, in sensitivity analysis which is often used to validate missing data handling processes by assessing robustness of imputations under different conditions, Bayesian model comparison criteria such as the deviance information criterion (DIC) can be used to choose the best model (Ma & Chen 2018; Van Buuren 2018). Further, there are ways to diagnose missing data mechanisms in a dataset such as discussed in Little (1988). These topics can be further explored.

There are also different analysis models tailored toward specific types of data that can be incorporated in Bayesian missing data handling processes, which we did not discuss in this paper. While this paper uses a structural equation model as an example, other models can be similarly constructed in JAGS to accommodate missing data. For example, longitudinal data can be analyzed using a multilevel model (Gelman & Hill 2006). If the response data are of mixed types from mixed populations, then mixture models may be appropriate (Rasmussen 1999). Social network data can also be used with methods such as the exponential random graph model and the latent space model (Hoff, Raftery, & Handcock 2002; Robins, Pattison, Kalish, & Lusher 2007). These models have Bayesian variants that can be applied to estimate parameters when missing data are present (Bürkner 2017; Koskinen, Robins, Wang, & Pattison 2013; Neal 1992).

Acknowledgment

The research reported here was supported, in whole or in part, by the Institute of Education Sciences, U.S. Department of Education, through grant R305D210023 to University of Notre Dame. The opinions expressed are those of the authors and do not represent the views of the Institute or the U.S. Department of Education.
References


Little, R. J. (1988). A test of missing completely at random for multivariate data with missing values. Journal of
A Appendix

A.1 Data Preparation

going set.seed(10)
data = read.csv("active.csv")
data = data[, c("WS_COR", "LS_COR", "LT_COR", "AGE", "EPT")]
data = na.omit(data)
data_500 <- data[sample(nrow(data),500),]
colnames(data_500) <- c("WS", "LS", "LT", "AGE", "EPT")

# MAR
miss <- rep(NA, nrow(data_500))
for (i in 1:nrow(data_500)){
  miss[i] <- rbinom(1, 1, data_500$AGE[i]*0.01-60*0.01)
}
data_500_mar <- data_500
data_500_mar[,"EPT"[miss==1] <- NA
summary(data_500_mar)

# MNAR selection
miss2 <- rep(NA, nrow(data_500))
for (i in 1:nrow(data_500)){
  miss2[i] <- rbinom(1, 1, data_500$EPT[i]*0.01)
}
data_500_mnar <- data_500
data_500_mnar[,"EPT"[miss2==1] <- NA
summary(data_500_mnar)

# MNAR pattern-mixture
miss3 <- rep(0, nrow(data_500))
miss3[(data_500$EPT>23) & (data_500$EPT<27)] <- 1
data_500_mnar_2 <- data_500
data_500_mnar_2[,"EPT"[miss3==1] <- NA
summary(data_500_mnar_2)

A.2 Ignorable Missingness Model

library(runjags)

ignorable <- "
model{
  # likelihood
  for (i in 1:N){
    EPT[i] ~ dnorm(mu.EPT[i], pre.EPT)
    mu.EPT[i] <- b0 + b1*REASON[i] + b2*AGE[i]
    WS[i] ~ dnorm(mu1[i], pre.WS)
    LT[i] ~ dnorm(mu2[i], pre.LT)
    LS[i] ~ dnorm(mu3[i], pre.LS)
  }
}
mu1[i] <- REASON[i] + k.WS
mu2[i] <- l.LT*REASON[i] + k.LT
mu3[i] <- l.LS*REASON[i] + k.LS
REASON[i] ~ dnorm(0, pre.REASON)
}

# priors
# regression model
b0 ~ dnorm(0, pre.b)
b1 ~ dnorm(0, pre.b)
b2 ~ dnorm(0, pre.b)
pre.b ~ dgamma(.001,.001)
pre.EPT ~ dgamma(0.001, 0.001)

# factor model
l.LT ~ dnorm(0, 0.001)
l.LS ~ dnorm(0, 0.001)
pre.WS ~ dgamma(0.001, 0.001)
pre.LT ~ dgamma(0.001, 0.001)
pre.LS ~ dgamma(0.001, 0.001)
k.WS ~ dnorm(0, 0.001)
k.LT ~ dnorm(0, 0.001)
k.LS ~ dnorm(0, 0.001)
pre.REASON ~ dgamma(0.001, 0.001)

# variances
var.EPT <- 1/pre.EPT
var.WS <- 1/pre.WS
var.LT <- 1/pre.LT
var.LS <- 1/pre.LS
var.REASON <- 1/pre.REASON
}

"data <- list(N = 500,
    EPT = data_500_mar$EPT,
    AGE = data_500_mar$AGE,
    WS = data_500_mar$WS,
    LT = data_500_mar$LT,
    LS = data_500_mar$LS)
inits <- list(list(b0=0, b1=0, b2=0, l.LT=0, l.LS=0,
    k.WS=0, k.LS=0, k.LT=0,
    pre.WS=0.1, pre.LS=0.1, pre.LT=0.1,
    pre.EPT=0.1, pre.REASON=0.1),}
out1 <- run.jags(model=ignorable, 
monitor=c("b0", "b1", "b2", 
"l.LT", "l.LS", 
"k.WS", "k.LT", "k.LS", 
"var.WS", "var.LT", 
"var.LS", "var.EPT", "var.REASON"), 
data=data, n.chains=2, 
inits=inits, method="simple", 
adapt=1000, burnin = 3000, 
sample=1000000, 
thin=1, 
keep.jags.files=FALSE, 
tempdir=TRUE)

out1
plot(out1)

A.3 Selection Model

selection<- "
model{
  # likelihood
  for (i in 1:N){
    R[i] ~ dbern(p[i])
    logit(p[i]) = a0 + a1*EPT[i]
    EPT[i] ~ dnorm(mu.EPT[i], pre.EPT)
    mu.EPT[i] <- b0 + b1*REASON[i] + b2*AGE[i]
    WS[i] ~ dnorm(mu1[i], pre.WS)
    LT[i] ~ dnorm(mu2[i], pre.LT)
    LS[i] ~ dnorm(mu3[i], pre.LS)
    mu1[i] <- REASON[i] + k.WS
    mu2[i] <- l.LT*REASON[i] + k.LT
    mu3[i] <- l.LS*REASON[i] + k.LS
    REASON[i] ~ dnorm(0, pre.REASON)
  }

  # priors
  # regression model
  b0 ~ dnorm(0, pre.b)
  b1 ~ dnorm(0, pre.b)
b2 ~ dnorm(0, pre.b)
a0 ~ dnorm(0, pre.a)
a1 ~ dnorm(0, pre.a)
pre.a ~ dgamma(.001,.001)
pre.b ~ dgamma(.001,.001)
pre.EPT ~ dgamma(0.001, 0.001)

# factor model
l.LT ~ dnorm(0, 0.001)
l.LS ~ dnorm(0, 0.001)
pre.WS ~ dgamma(0.001, 0.001)
pre.LT ~ dgamma(0.001, 0.001)
pre.LS ~ dgamma(0.001, 0.001)
k.WS ~ dnorm(0, 0.001)
k.LT ~ dnorm(0, 0.001)
k.LS ~ dnorm(0, 0.001)
pre.REASON ~ dgamma(0.001, 0.001)

# variances
var.EPT <- 1/pre.EPT
var.WS <- 1/pre.WS
var.LT <- 1/pre.LT
var.LS <- 1/pre.LS
var.REASON <- 1/pre.REASON

"data <- list(N = 500,
  EPT = data_500_mnar$EPT,
  AGE = data_500_mnar$AGE,
  WS = data_500_mnar$WS,
  LT = data_500_mnar$LT,
  LS = data_500_mnar$LS,
  R = is.na(data_500_mnar$EPT)*1)

inits <- list(list(b0=0, b1=0, b2=0, l.LT=0, l.LS=0,
  k.WS=0, k.LS=0, k.LT=0,
  pre.WS=0.1, pre.LS=0.1, pre.LT=0.1,
  pre.EPT=0.1, pre.REASON=0.1,
  a0=0.1, a1=0.1),
  list(b0=1, b1=1, b2=1, l.LT=1, l.LS=1,
   k.WS=1, k.LS=1, k.LT=1,
   pre.WS=0.2, pre.LS=0.2, pre.LT=0.2,
   pre.EPT=0.2, pre.REASON=0.2,
   a0=0.2, a1=0.2))
out1 <- run.jags(model=selection,
   monitor=c("b0", "b1", "b2",
   "a0", "a1",
   "l.LT", "l.LS", "k.WS",
   "k.LT", "k.LS",
   "var.WS", "var.LT",
   "var.LS", "var.EPT", "var.REASON"),
   data=data, n.chains=2,
   inits=inits, method="simple",
   adapt=1000, burnin = 3000,
   sample=1000000,
   thin=1,
   keep.jags.files=FALSE,
   tempdir=TRUE)
out1
plot(out1)

A.4 Pattern-Mixture Model

pmm <- "
model{
  # likelihood
  for (i in 1:N){
    EPT[i] ~ dnorm(mu.EPT[i], pre.EPT)
    mu.EPT[i] <- b0[R[i]+1] + b1[R[i]+1]*REASON[i] + b2*AGE[i]
    WS[i] ~ dnorm(mu1[i], pre.WS)
    LT[i] ~ dnorm(mu2[i], pre.LT)
    LS[i] ~ dnorm(mu3[i], pre.LS)
    mu1[i] <- REASON[i] + k.WS
    mu2[i] <- l.LT*REASON[i] + k.LT
    mu3[i] <- l.LS*REASON[i] + k.LS
    REASON[i] ~ dnorm(0, pre.REASON)
  }
  # priors
  # regression model
  b0[1] ~ dnorm(0, pre.b) # non-missing
  b0[2] ~ dnorm(b0[1]+0.5, pre.b) # missing
  b1[1] ~ dnorm(0, pre.b) # non-missing
  b1[2] ~ dnorm(b1[1]-0.5, pre.b) # missing
  b2 ~ dnorm(0, pre.b)
  pre.b ~ dgamma(0.001, 0.001)
  pre.EPT ~ dgamma(0.001, 0.001)"
# factor model
l.LT ~ dnorm(0, 0.001)
l.LS ~ dnorm(0, 0.001)
pre.WS ~ dgamma(0.001, 0.001)
pre.LT ~ dgamma(0.001, 0.001)
pre.LS ~ dgamma(0.001, 0.001)
k.WS ~ dnorm(0, 0.001)
k.LT ~ dnorm(0, 0.001)
k.LS ~ dnorm(0, 0.001)
pre.REASON ~ dgamma(0.001, 0.001)

# variances
var.EPT <- 1/pre.EPT
var.WS <- 1/pre.WS
var.LT <- 1/pre.LT
var.LS <- 1/pre.LS
var.REASON <- 1/pre.REASON

data <- list(N = 500,
  EPT = data_500_mnar_2$EPT,
  AGE = data_500_mnar_2$AGE,
  WS = data_500_mnar_2$WS,
  LT = data_500_mnar_2$LT,
  LS = data_500_mnar_2$LS,
  R = is.na(data_500_mnar_2$EPT)*1)
inits <- list(list(b0=c(0,1), b1=c(1,0), b2=0,
  l.LT=0, l.LS=0, k.WS=0,
  k.LS=0, k.LT=0,
  pre.WS=0.1, pre.LS=0.1, pre.LT=0.1,
  pre.EPT=0.1, pre.REASON=0.1),
  list(b0=c(1,2), b1=c(2,1), b2=1,
  l.LT=1, l.LS=1, k.WS=1,
  k.LS=1, k.LT=1,
  pre.WS=0.2, pre.LS=0.2, pre.LT=0.2,
  pre.EPT=0.2, pre.REASON=0.2))
out1 <- run.jags(model=pmm,
  monitor=c("b0", "b1", "b2",
  "l.LT", "l.LS",
  "k.WS", "k.LT", "k.LS",
  "var.WS", "var.LT",
  "var.REASON","var.EPT", "pre.WS", "pre.LT", "pre.LS", "pre.EPT", "pre.REASON"))
"var.LS", "var.EPT", "var.REASON"),
data=data, n.chains=2,
init=init, method="simple",
adapt=1000, burnin = 3000,
sample=1000000,
thin=1,
keep.jags.files=FALSE,
tempdir=TRUE)
out1
plot(out1)