Hierarchical Bayesian Model Selection

The Role of Nutritional Supplementation on Betaine Concentration in Pregnant Pakistani Women

Nicholas Weaver

Introduction

- Hierarchical Bayesian Models occur when the parameter(s) of a prior distribution has/have a prior distribution
  - Common in correlated responses
- Betaine has been connected to pregnancy outcomes such as infant length at birth.
  - Nutritional supplements for malnourished pregnant women may have an impact on the concentration of Betaine in the mother after 34 weeks gestation.

The Data

- 31 pregnant Pakistani women
  - 13 on NS since preconception (Arm 1)
  - 18 on NS since 12 Weeks gestation (Arm 2)
- Women from 13 different locations in Pakistan (clusters, denoted with subscript i)
  - Implies correlated responses

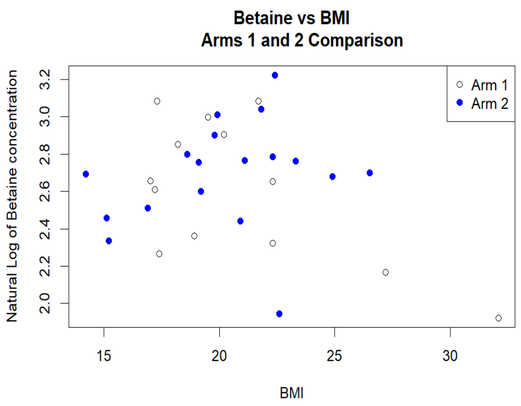

Figure 1. Betaine Scatterplot by BMI: The natural log of betaine concentration for the observed Pakistani women. Values from Arm 1 and Arm 2 are signified by color (white and blue, respectively). Future work will consider the role of BMI (the x-axis).

Acknowledgments

- Dr. Joshua French and his course MATH 7393
- Dr. Audrey Hendricks, Dr. Stephanie Gilley, and Dr. Sarah Borengasser for guiding this study
- The Women First Research Group led by Dr. Michael Hambidge and Dr. Nancy Krebs for providing the data and producing the overall study

Model Designs

$\boldsymbol{Y_{i,j}}$: Betaine concentration for the j^{th} person from the i^{th} cluster.

$\boldsymbol{\beta_0}$: parameter for the mean metabolite concentration of Women from Arm 2

$\boldsymbol{\beta_1}$: parameter for the difference in mean metabolite concentration between Arm1 and Arm 2 Women

Model with No Random Effect (NRE)

Sampling Distribution:

$$\ln(Y_{i,j}) \sim N(\mu_{i,j}, \sigma^2)\ , \mu_{i,j} = \beta_0 + \beta_1 X_{i,j} + \beta_i$$

Prior Distributions:

$$\beta_k \sim N(0, 1{,}000)\ , \mathrm{k} = 0{,}1, \dots, 13$$
$$\sigma^2 \sim IG(0.001, 0.001)$$

Model with Random Effect for Site (RES)

Sampling Distribution:

$$\ln(Y_{i,j}) \sim N(\mu_{i,j} + \alpha_i, \sigma_0^2)\ , \mu_{i,j} = \beta_0 + \beta_1 X_{i,j}$$

Prior Distributions:

$$\beta_k \sim N(0, 1{,}000)\ , \mathrm{k} = 0, 1$$
$$\alpha_i \sim N(0, \sigma_1^2)$$
$$\sigma_l^2 \sim IG(0.001, 0.001),\ l = 0, 1$$

Model with Random Effect for Site and Woman (RESW)

Sampling Distribution:

$$\ln(Y_{i,j,k}) \sim N(\mu_{i,j} + \alpha_i + \gamma_k, \sigma_0^2)\ , \mu_{i,j} = \beta_0 + \beta_1 X_{i,j}$$

Prior Distributions:

$$\beta_m \sim N(0, 1{,}000)\ , \mathrm{m} = 0, 1$$
$$\alpha_i \sim N(0, \sigma_1^2)$$
$$\gamma_k \sim N(0, \sigma_2^2)$$
$$\sigma_l^2 \sim IG(0.001, 0.001),\ l = 0, 1, 2$$

Where k is the k^{th} observation

Model Analysis

Model	Variance Breakdown	Mean	SD	2.5 Percentile	97.5 Percentile
NRE	σ^2	0.1094	0.0428	0.0544609	0.2171397
	Total Variance	0.1094	0.0428	0.0544609	0.2171397
RES	σ^2	0.0993	0.0316	0.0535146	0.1758947
	σ_1^2	0.0219	0.0259	0.0008010	0.0903016
	Total Variance	0.1212	0.0375	0.0696655	0.2118633
RESW	σ^2	0.0510	0.0431	0.0009505	0.1449823
	σ_1^2	0.0216	0.0260	0.0007846	0.0897476
	σ_2^2	0.0508	0.0431	0.0008985	0.1452992
	Total Variance	0.1234	0.0384	0.0706971	0.2162895

Table 1. Random Effects: Posterior values for distributions of the variance components for each model.

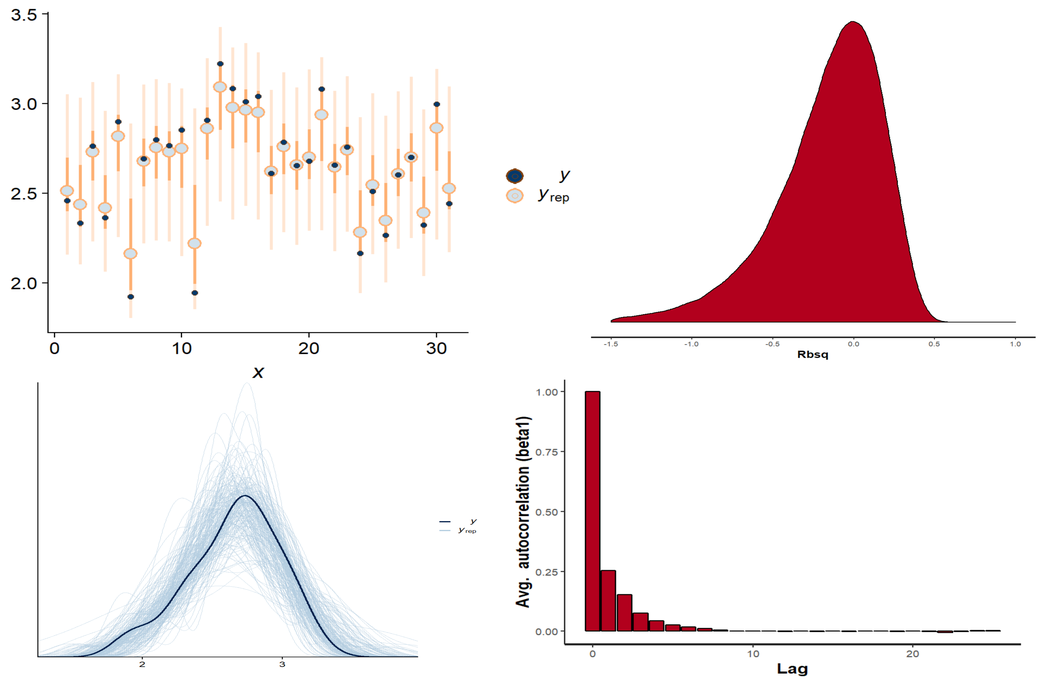

Figure 3. Model verification: (**TL**) 50% (dark orange), 90% (light orange) bands and mean estimate (gray circle) for the replicated response value at each observed response (dark blue circle). Calculated from 10,000 replicated data sets. (**TR**) Density of the Bayesian R-squared value for the RESW model with mean estimate -0.171. (**BL**) Overlay of 200 replicated data densities generated from the posterior distributions of the RESW model. (**BR**) Average autocorrelation for $\boldsymbol{\beta_1}$ in RESW model.

Model Selection

Model	LOOIC
NRE	43.382
RES	23.951
RESW	11.678

Table 2. Selection by LOOIC: Leave-one-out-information-criterion (LOOIC) recommends the model with the smallest value (RESW).

Results

Parameter	Posterior Mean	95% CPI
β_1	-0.106	(-0.349, 0.138)

Table 3. Arm Parameter Results: The posterior mean and 95% central posterior interval for the distribution of the β_1 parameter. Results are from the RESW model, as determined through model selection.

Conclusion/Future Work

- Arm 1 and Arm 2 do not appear to have significant differences in terms of metabolic concentration among the Pakistani women.
- The final model (RESW) has poor predictive ability
- Consider the interaction of BMI and Arm type to potentially strengthen the model