Mixed-effects example
Adapted from *Applied Linear Statistical Methods* by Neter et al.

Sheffield Foods Company markets a variety of dairy products, including milk, ice cream, and yogurt. Recently, the company received a complaint from a government agency that the actual levels of milk fat in its yogurt exceeded the labeled amount. Company personnel were concerned that the government’s laboratory method for measuring fat content in yogurt might be unreliable because it is primarily designed for use with milk and ice cream. To study the reliability of Sheffield’s and the government’s laboratory methods, a small inter-laboratory study was carried out. Four testing laboratories were randomly selected from the population of laboratories in the United States. Each laboratory was sent 12 samples of yogurt, with instructions to evaluate six of the samples using the government’s method and six by the company’s method. The yogurt had been mixed under carefully controlled conditions and the fat content of each sample was known to be 3.0 percent.

Due to technical difficulties with the government method, none of the laboratories was able to obtain fat content determinations for all of the six samples assigned to that method in the time available. The data are presented in the following table:

<table>
<thead>
<tr>
<th>Laboratories</th>
<th>Method</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Government</td>
<td>5.19</td>
<td>4.09</td>
<td>4.62</td>
<td>3.71</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5.09</td>
<td>3.00</td>
<td>4.32</td>
<td>3.86</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.75</td>
<td>4.35</td>
<td>3.79</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4.04</td>
<td>4.59</td>
<td>3.63</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4.06</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sheffield</td>
<td>3.26</td>
<td>3.02</td>
<td>3.08</td>
<td>2.98</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.38</td>
<td>3.32</td>
<td>2.95</td>
<td>2.89</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.24</td>
<td>2.83</td>
<td>2.98</td>
<td>2.75</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.41</td>
<td>2.96</td>
<td>2.74</td>
<td>3.04</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.35</td>
<td>3.23</td>
<td>3.07</td>
<td>2.88</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.04</td>
<td>3.07</td>
<td>2.70</td>
<td>3.20</td>
<td></td>
</tr>
</tbody>
</table>

Model:

\[ y_{ijk} = \mu + \gamma_i + \beta_j + (\alpha \beta)_{ij} + \epsilon_{ijk} \]

- Fixed effect is measurement method with two levels.
- Random effect is laboratories with four levels.
- \( \beta_j \), \((\alpha \beta)_{ij}\), and \( \epsilon_{ijk} \) are independent \( \forall i, j, k \)
- First level of hierarchy:
  1. \( \beta_j \sim N(0, \sigma^2_{\beta}) \), \( j = 1, \ldots, 4 \)
  2. \((\alpha \beta)_{ij} \sim N(0, \sigma^2_{\alpha \beta}) \), \( i = 1, 2; j = 1, \ldots, 4 \)
  3. \( \epsilon_{ijk} \sim N(0, \sigma^2) \), \( i = 1, 2; j = 1, \ldots, 4; k = 1, \ldots, n_j \)
  4. \( \alpha \sim N(3, [0.0001]^{-1}) \)
- Second level of hierarchy:
  1. \( \sigma^2_{\beta} \sim \text{Inv-Gamma}(0.0001, 0.0001) \)
  2. \( \sigma^2_{\alpha \beta} \sim \text{Inv-Gamma}(0.0001, 0.0001) \)
  3. \( \sigma^2 \sim \text{Inv-Gamma}(0.0001, 0.0001) \)

Analysis:

- Calculations were carried out using WinBUGS. See attached program.
- Initial values for \( \sigma^2 \), \( \sigma^2_{\beta} \), and \( \sigma^2_{\alpha \beta} \) were set at their maximum likelihood estimates. All other initial values were randomly generated.
- To reproduce the results shown here set the WinBUGS seed to 939877734.
Numerical results are based on 5,000 iterations after a burn-in period of 5,000 iterations.

**Results**

- 95% posterior credible interval for $\sigma_{\alpha\beta}^2$: (0.001, 0.098). Hence, it seems that there are no interaction effects between measurement methods and laboratories.

- 95% posterior credible interval for $\sigma_{\beta}^2/(\sigma^2 + \sigma_{\beta}^2 + \sigma_{\alpha\beta}^2/2)$: (0.196, 0.946), this, together with the fact that the posterior median is 0.6308 and the strong asymmetry of its posterior distribution, reflects that most of the variation on the measurements is due to the laboratories instead of the measurement method.

- 95% posterior credible interval for $\alpha_1 - \alpha_2$: (1.014, 1.376). Hence, there is evidence that the government’s method for fat determination produces consistently higher values than the Sheffield’s method. The posterior evidence shows that the government’s method tends to obtain measurements that are at least 1% higher than the Sheffield’s method.
Measurement method

WinBUGS code

```r
model{ for (i in 1:39){
  y[i] ~ dnorm(mu[i],tau)
  mu[i] <- alpha[1]*x1[i]+alpha[2]*x2[i]+ beta[lab[i]]+ab[i,lab[i]]
}
# Higher level definitions
for (j in 1:4) { beta[j] ~ dnorm(0,tau.b)
  for (k in 1:39) {ab[k,j] ~ dnorm(0,tau.ab) }}
# Priors for fixed effects
alpha[1] ~ dnorm(3,0.0001)
alpha[2] ~ dnorm(3,0.0001)
# Priors for random terms
tau ~ dgamma(0.001,0.001)
tau.b ~ dgamma(0.001,0.001)
tau.ab ~ dgamma(0.001,0.001)
#Measures of interest
diff.a <- alpha[1]-alpha[2]
s2 <- 1/tau
s2.b <- 1/tau.b
s2.ab <- 1/tau.ab
var.total <- s2.b + s2.ab/2 + s2
var.lab<- s2.b/var.total
}
#Data
list(y = c(5.19, 5.09, 4.09, 3.99, 3.75, 4.04, 4.06, 4.62, 4.32, 4.35, 4.59, 3.71, 3.86, 3.79, 3.63, 3.26, 3.48, 3.24, 3.41, 3.35, 3.04, 3.02, 3.32, 2.83, 2.96, 3.23, 3.07, 3.08, 2.95, 2.98, 2.74, 3.07, 2.7, 2.98, 2.89, 2.75, 3.04, 2.88, 3.2),
```

5

6
# Initial values are set close to the MLE estimates
list( tau = 44, tau.b = 11, tau.ab = 13)