36-463 / 36-663: Multilevel & Hierarchical Models
HW06 Solution

October 30, 2016

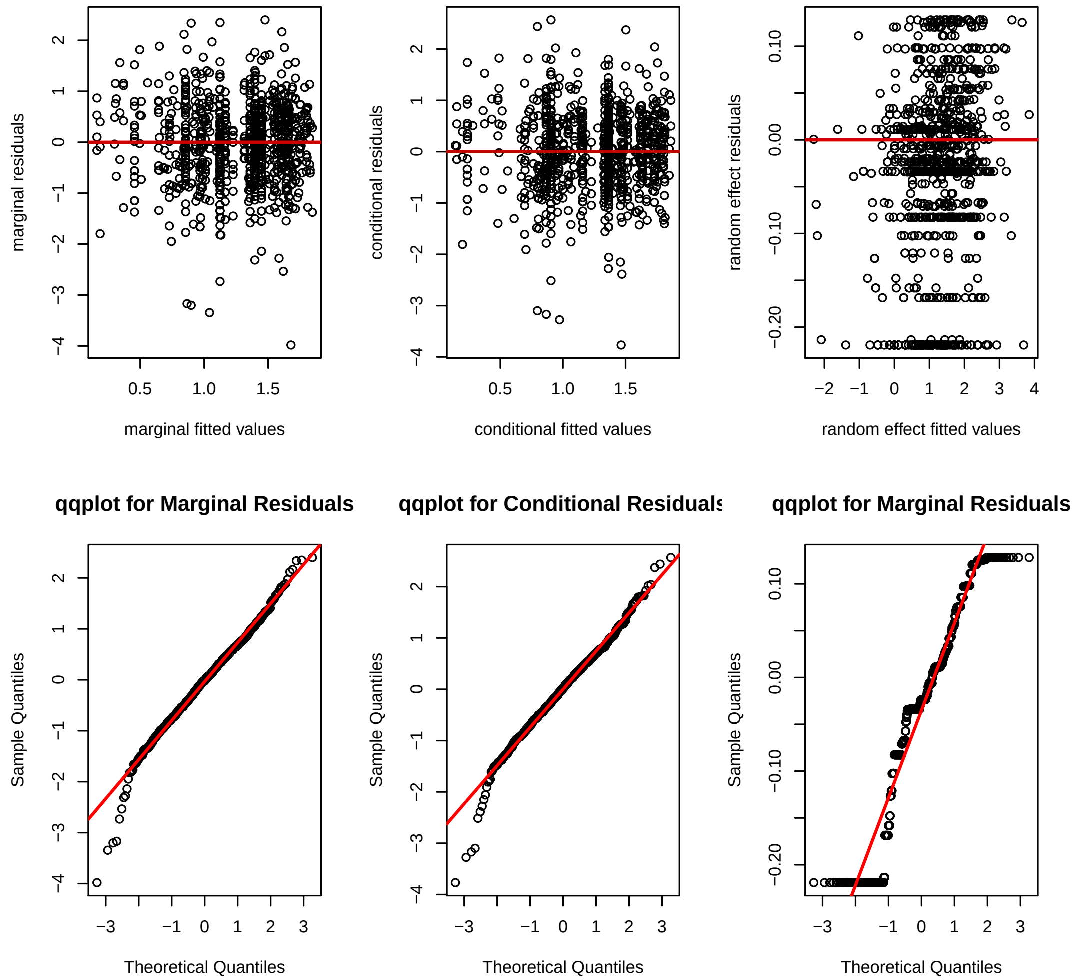
Question 1

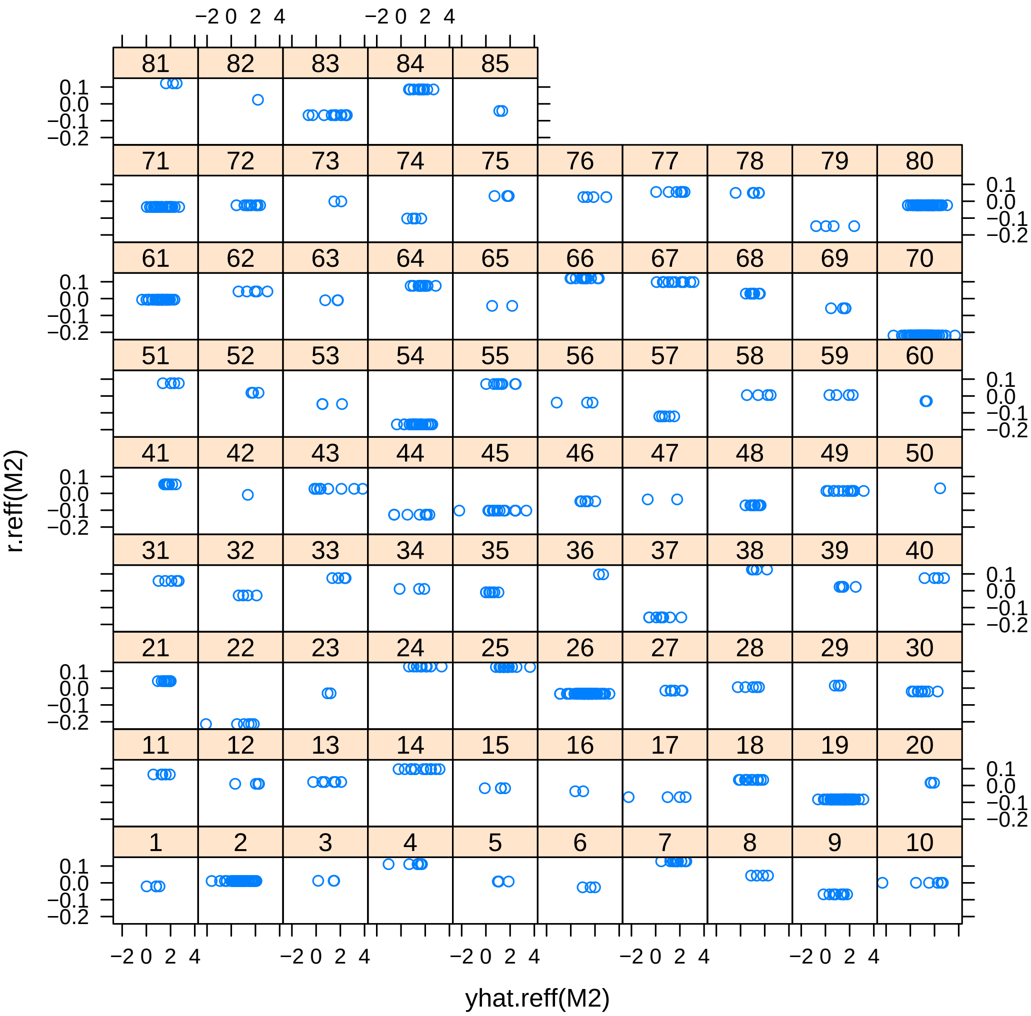
```
library(lme4)
library(arm)
mn.radon <- read.table("mn-radon.txt")
attach(mn.radon)

n <- length(radon)
j <- length(unique(county))
y <- log.radon
x <- floor
u.full <- log.uranium
M0 <- lmer( y ~ 1 + (1|county))
M1 <- lmer(y ~ x + (1|county))
M2 <- lmer(y ~ x + u.full + (1|county))
M3 <- lmer(y ~ x + u.full +(1+ x|county))

anova(M0,M1,M2,M3)
Data:
Models:
M0: y ~ 1 + (1 | county)
M1: y ~ x + (1 | county)
M2: y ~ x + u.full + (1 | county)
M3: y ~ x + u.full + (1 + x | county)
   Df    AIC    BIC  logLik   Chisq Chi Df Pr(>Chisq)
M0  3 2261.2 2275.7 -1127.6
M1  4 2171.7 2190.9 -1081.8 91.5876      1  < 2.2e-16 ***
M2  5 2132.9 2157.0 -1061.4 40.7947      1  1.691e-10 ***
M3  7 2131.7 2165.4 -1058.8  5.2002      2    0.07426 .

display(M0)
number of obs: 919, groups: county, 85
AIC = 2265.4, DIC = 2251
deviance = 2255.2
display(M1)
number of obs: 919, groups: county, 85
AIC = 2179.3, DIC = 2156
deviance = 2163.7
display(M2)
```

```
number of obs: 919, groups: county, 85
AIC = 2144.2, DIC = 2111.5
deviance = 2122.9
display(M3)
number of obs: 919, groups: county, 85
AIC = 2142.6, DIC = 2106.7
deviance = 2117.7

code for producing all the residuals plot for M2, M3 is similar:
par(mfrow = c(2,3))
plot(yhat.marg(M2),r.marg(M2),xlab="marginal fitted values",ylab="marginal residuals")
abline(h=0,lwd = 2, col = "red3")
plot(yhat.cond(M2),r.cond(M2),xlab="conditional fitted values",ylab="conditional residuals")
abline(h=0,lwd = 2, col = "red3")
plot(yhat.reff(M2),r.reff(M2),xlab="random effect fitted values",ylab="random effect residuals")
abline(h=0,lwd = 2, col = "red3")
n2<-qqnorm(r.marg(M2), main="qqplot for Marginal Residuals")
abline(lm(y~x,data=n2),col="red",lwd=2)
n2<-qqnorm(r.cond(M2), main="qqplot for Conditional Residuals")
abline(lm(y~x,data=n2),col="red",lwd=2)
n2<-qqnorm(r.reff(M2), main="qqplot for Marginal Residuals")
abline(lm(y~x,data=n2),col="red",lwd=2)
xyplot(r.reff(M2)~yhat.reff(M2)|as.factor(county))
```

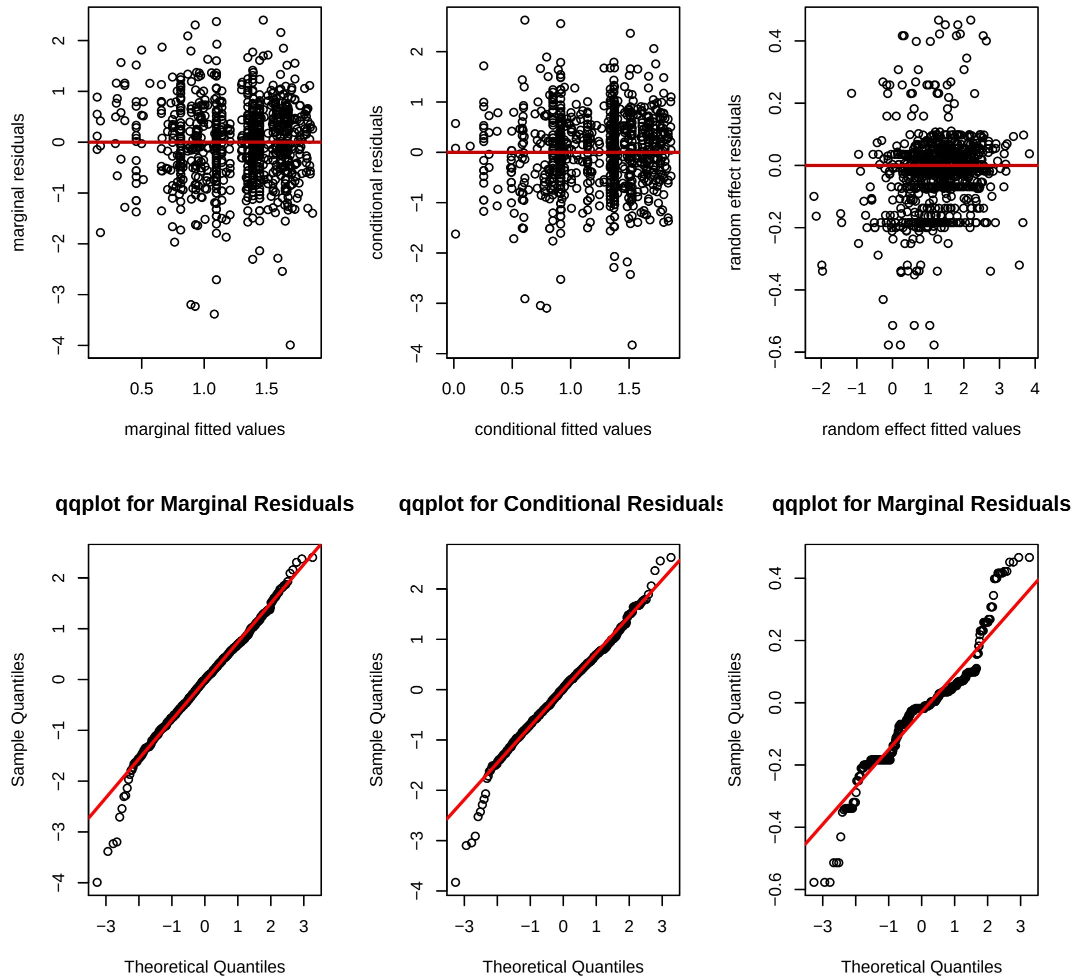
We can see that based on AIC, M2 and M3 seem equally good, and better than M1. Based on BIC, M2 seems best; based on DIC, M3 seems best.

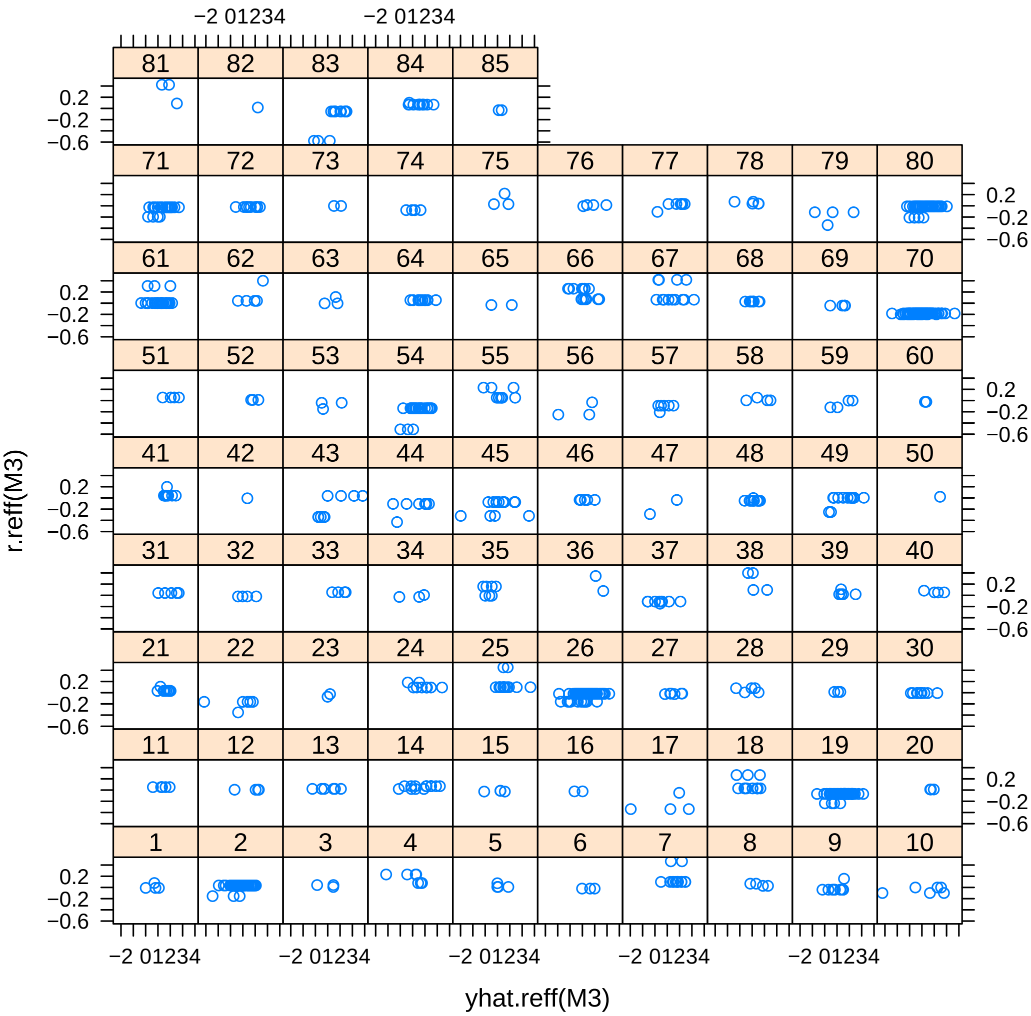
Plot for M2:


marginal residuals
marginal fitted values
conditional residuals
conditional fitted values
random effect residuals
random effect fitted values
qqplot for Marginal Residuals
Sample Quantiles
Theoretical Quantiles
qqplot for Conditional Residuals
Sample Quantiles
Theoretical Quantiles
qqplot for Marginal Residuals
Sample Quantiles
Theoretical Quantiles



Plot for M3:


marginal residuals
marginal fitted values
conditional residuals
conditional fitted values
random effect residuals
random effect fitted values
qqplot for Marginal Residuals
qqplot for Conditional Residuals
qqplot for Marginal Residuals
Sample Quantiles
Theoretical Quantiles
Sample Quantiles
Theoretical Quantiles
Sample Quantiles
Theoretical Quantiles



All residual plots show similar properties between M2 and M3.

One could use a likelihood ratio test to compare M1 with M2, or M1 with M3, but not M2 with M3 (because we get M2 from M3 by setting one of the variance components to zero, which is at the edge of the parameter space where the usual χ^2 theory for likelihood ratio tests fails). To compare M2 and M3 directly, we could try the simulation-based "exact likelihood ratio test" (which uses simulation to find the true distribution of the test statistic under the null hypothesis [it is no longer χ^2]).

For example, if we want to compare M2 and M3, we can compare their deviance and degree of freedom and use chi-square test statistics to see if they are significantly different etc. For example with `RLRsim` we would get:

```
> library(RLRsim)
>
> M2 <- lmer(y ~ x + u.full + (1 | county) )          # the "null hypothesis model"
> M3 <- lmer(y ~ x + u.full + (1|county) + (0 + x | county))   # the "alternative" model
>
> M.helper <- lmer(y~ x + u.full + (0 + x|county))   # helper model needed for the simulation
>
```

```
> exactRLRT(m=M.helper,mA=M3,m0=M2)

        simulated finite sample distribution of RLRT.

        (p-value based on 10000 simulated values)

data:
RLRT = 5.3938, p-value = 0.0093
```

So it appears that M3 really is preferred.

Note that there is a little sleight-of-hand here. The original M3 was `y ~ x + u.full + (1 + x|county)`, but `exactRLRT` cannot handle correlated random effects, so we fitted M3 as `y ~ x + u.full + (1|county) + (0 + x|county)` instead, to make the random effects independent. This somewhat changes the AIC, BIC, and DIC comparisons also:

```
anova(M0,M1,M2,M3)
   Df    AIC    BIC  logLik deviance   Chisq Chi Df Pr(>Chisq)
M0  3 2261.2 2275.7 -1127.6   2255.2
M1  4 2171.7 2190.9 -1081.8   2163.7 91.5870      1  < 2.2e-16 ***
M2  5 2132.8 2156.9 -1061.4   2122.8 40.8337      1  1.658e-10 ***
M3  6 2129.8 2158.8 -1058.9   2117.8  5.0104      1     0.0252 *
display(M2)
number of obs: 919, groups: county, 85
AIC = 2144.2, DIC = 2111.5
deviance = 2122.9
display(M3)
number of obs: 919, groups: county, 85
AIC = 2140.8, DIC = 2106.8
deviance = 2117.8
```

although the conclusions are basically the same as before: M3 is preferred by AIC and DIC and M2 is preferred by BIC.

There is no exact simulation test available in R for testing correlated random effects, but based on the evidence that we have gathered so far, M3 (with or without the correlation between the random slope and random intercept on `x`) seems the slightly better model.

Question 2

(a)

$$\begin{aligned} y_i &= \beta_0 + \eta_{j[i]} + \varepsilon_i, \\ \varepsilon_i &\sim N(0, \sigma^2) \\ \eta_j &\sim N(0, \tau^2) \\ & i = 1, \ldots, n; j = 1, \ldots, m \end{aligned}$$

i. This is in **variance components** form.

ii. The other two representations are:
multilevel model:

$$\begin{aligned} y_i &= \alpha_{0j[i]} + \varepsilon_i \\ \alpha_{0j} &= \beta_0 + \eta_{0j} \\ \varepsilon_i &\sim N(0, \sigma^2) \\ \eta_{0j} &\sim N(0, \tau^2) \\ & i = 1, \ldots, n; j = 1, \ldots, m \end{aligned}$$

Heirarchical:

$$\begin{aligned} y_i &\sim N(\alpha_{0j[i]}, \sigma^2) \\ \alpha_{0j} &\sim N(\beta_0, \tau^2) \\ & \quad i = 1, \ldots, n; j = 1, \ldots, m \end{aligned}$$

iii. The lmer model notation is

```
y ~ 1 + (1|group)
```

where "group" is the thing that varies with j above.

(b)

$$\begin{aligned} y_i &\sim N(\alpha_{0j[i]} + \alpha_1 x_i, \sigma^2) \\ \alpha_{0j} &\sim N(\beta_0 + \beta_1 z_j, \tau^2) \\ & \quad i = 1, \ldots, n; j = 1, \ldots, m \end{aligned}$$

i. This is in **hierarchical** form.

ii. The other two forms are:

multilevel:

$$\begin{aligned} y_i &= \alpha_{0j[i]} + \alpha_1 x_i + \varepsilon_i \\ \alpha_{0j} &= \beta_0 + \beta_1 z_j + \eta_{0j} \\ \varepsilon_i &\sim N(0, \sigma^2) \\ \eta_{0j} &\sim N(0, \tau^2) \\ & \quad i = 1, \ldots, n; j = 1, \ldots, m \end{aligned}$$

variance components:

$$\begin{aligned} y_i &= \beta_0 + \beta_1 z_{j[i]} + \alpha_1 x_i + \eta_{0j[i]} + \varepsilon_i \\ \varepsilon_i &\sim N(0, \sigma^2) \\ \eta_{0j} &\sim N(0, \tau^2) \\ & \quad i = 1, \ldots, n; j = 1, \ldots, m \end{aligned}$$

iii. The lmer syntax is

```
   y ~ z + x + (1|group)
or
   y ~ 1 + z + x + (1|group)     [the intercept is included tacitly...]
```

where "group" is the thing that varies with j above.

(c)

$$\begin{aligned} y_i &= \beta_{00} + \beta_{10} x_i + \beta_{11} x_i z_{j[i]} + \eta_{0j[i]} + \eta_{1j[i]} x_i + \varepsilon_i \\ \varepsilon_i &\sim N(0, \sigma^2) \\ \eta_{0j} &\sim N(0, \tau_0^2) \\ \eta_{1j} &\sim N(0, \tau_1^2) \\ & \quad i = 1, \ldots, n; j = 1, \ldots, m \end{aligned}$$

i. This is in **variance components** form.

ii. The other two forms are

multilevel model:

$$
\begin{aligned}
y_i &= \alpha_{0j[i]} + \alpha_{1j[i]} x_i + \varepsilon_i \\
\alpha_{0j} &= \beta_{00} + \eta_{0j} \\
\alpha_{1j} &= \beta_{10} + \beta_{11} z_j + \eta_{1j} \\
\eta_{0j} &\sim N(0, \tau_0^2) \\
\eta_{1j} &\sim N(0, \tau_1^2) \\
& \quad i = 1, \ldots, n; j = 1, \ldots, m
\end{aligned}
$$

Note: You could equally well remove the term "β_{10}" from the equation for α_{1j} and put a term "$\beta_{10} x_i$" in the equation for y_i... But the way I have written it has the possible intuitive advantage of having an intercept in the linear model at each "level".

hierarchical:

Following the multilevel model as I have written it above,

$$
\begin{aligned}
y_i &\sim N(\alpha_{0j[i]} + \alpha_{1j[i]} x_i, \sigma^2) \\
\alpha_{0j} &\sim N(\beta_{00}, \tau_0^2) \\
\alpha_{1j} &\sim N(\beta_{10} + \beta_{11} z_j, \tau_1^2) \\
& \quad i = 1, \ldots, n; j = 1, \ldots, m
\end{aligned}
$$

iii. The lmer model notation is:

```
y ~ x + x:z + (x|group)
```

or

```
y ~ 1 + x + x:z + (1 + x | group)   [the 1's are included tacitly...]
```

where "group" is the thing that varies with j above.

Note: This one was a little trickier than I intended, because I did not include a linear fixed-effects term in z in the variance-components model. If I had included the z term the lmer model would have been the somewhat more conventional `y ~ x + x*z + (x|group)`).

(d)

$$
\begin{aligned}
y_i &= \alpha_{0j[i]} + \alpha_{1j[i]} x_i + \varepsilon_i \\
\alpha_{0j} &= \beta_{00} + \beta_{01} z_j + \eta_{0j} \\
\alpha_{1j} &= \beta_{10} + \beta_{11} w_j + \eta_{1j} \\
\varepsilon_i &\sim N(0, \sigma^2) \\
\eta_{0j} &\sim N(0, \tau_0^2) \\
\eta_{1j} &\sim N(0, \tau_1^2) \\
& \quad i = 1, \ldots, n; j = 1, \ldots, m
\end{aligned}
$$

i. This model is in **multilevel** form.

ii. The other forms are:

hierarchical:

$$
\begin{aligned}
y_i &\sim N(\alpha_{0j[i]} + \alpha_{1j[i]} x_i, \sigma^2) \\
\alpha_{0j} &\sim N(\beta_{00} + \beta_{01} z_j, \tau_0^2) \\
\alpha_{1j} &\sim N(\beta_{10} + \beta_{11} w_j, \tau_1^2) \\
& \quad i = 1, \ldots, n; j = 1, \ldots, m
\end{aligned}
$$

variance components:

$$
\begin{aligned}
y_i &= \beta_{00} + \beta_{01} z_{j[i]} + \beta_{10} x_i + \beta_{11} x_i w_{j[i]} + \eta_{0j[i]} + \eta_{1j[i]} x_i + \varepsilon_i \\
\varepsilon_i &\sim N(0, \sigma^2) \\
\eta_{0j} &\sim N(0, \tau_0^2) \\
\eta_{1j} &\sim N(0, \tau_1^2) \\
& \quad i = 1, \ldots, n; j = 1, \ldots, m
\end{aligned}
$$

iii. The lmer model notation is:

```
   y ~ z + x + x:w + (x|group)
or
   y ~ 1 + z + x + x:w + (1+x|group)   [the 1's are added tacitly...]
```

where "group" is the thing that varies with j above.