Homework 7

Xinyi Lin

4/14/2019

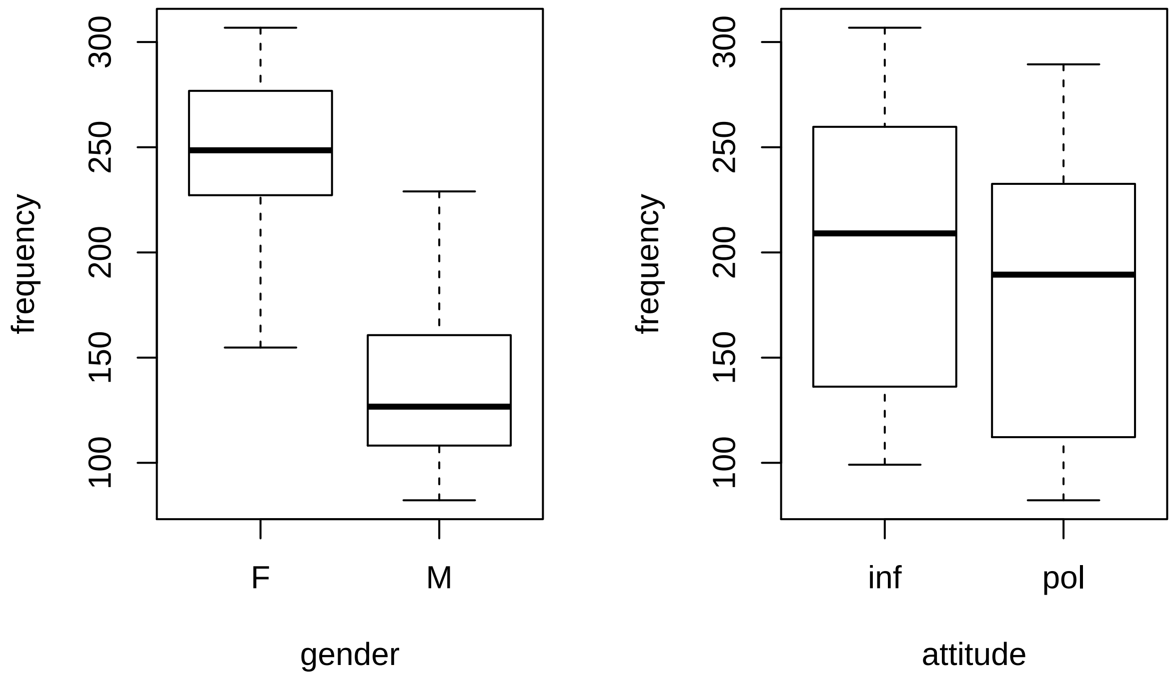
```
library(nlme)
```

Import data

```
data = read.csv("./politeness_data.csv")
```

Question 1

```
par(mfrow=c(1,2))
boxplot(data$frequency~data$gender, xlab = "gender", ylab = "frequency")
boxplot(data$frequency~data$attitude, xlab = "attitude", ylab = "frequency")
```



According to boxplots, we can find that overall 1) pitches of women is significantly higher than pitches of men; 2) pitch of the inf level is slightly higher than the pol level.

Question 2

Fit model 1.

```
lmm1 <- lme(frequency ~ gender + attitude, random = ~1 | subject,
            data = data, method='REML')
summary (lmm1)
```

```
## Linear mixed-effects model fit by REML
##  Data: data
```

```
##         AIC      BIC    logLik
##   806.0805 818.0527 -398.0402
##
## Random effects:
##  Formula: ~1 | subject
##         (Intercept) Residual
## StdDev:    24.45803 29.11537
##
## Fixed effects: frequency ~ gender + attitude
##                 Value Std.Error DF   t-value p-value
## (Intercept)  256.98690 15.154986 77 16.957251  0.0000
## genderM     -108.79762 20.956235  4 -5.191659  0.0066
## attitudepol  -20.00238  6.353495 77 -3.148248  0.0023
##  Correlation:
##             (Intr) gendrM
## genderM     -0.691
## attitudepol -0.210  0.000
##
## Standardized Within-Group Residuals:
##        Min         Q1        Med         Q3        Max
## -2.3564422 -0.5658319 -0.2011979  0.4617895  3.2997610
##
## Number of Observations: 84
## Number of Groups: 6
```

As

$$Y_{ij} = \beta_0 + X_{ij}^T\beta + b_i + \epsilon_{ij}$$

where $b_i \sim N(0, \sigma_b^2)$ and $\epsilon_{ij} \sim N(0, \sigma^2)$.

$$var(Y_i) = var[(\beta_1 + b_i) + X_{ij}^T\beta + b_i + \epsilon_{ij}] = \sigma_b^2 + \sigma^2$$

$$cor(Y_{ij}, Y_{ik}) = \sigma_b^2$$

```r
#VarCorr(lmm1)
sigma_b = as.numeric(VarCorr(lmm1)[1,2])^2
sigma = as.numeric(VarCorr(lmm1)[2,2])^2
sigma + sigma_b # variance of yij
```

```
## [1] 1445.9
```

```r
sigma_b # covariance
```

```
## [1] 598.1952
```

So the covariance matrix for a subject Y_i is:

$$\begin{bmatrix} 1445.9 & 598.2 & \dots & 598.2 \\ 598.2 & 1445.9 & \dots & 598.2 \\ \vdots & \vdots & \ddots & \vdots \\ 598.2 & 598.2 & \dots & 1445.9 \end{bmatrix}_{14\times 14}$$

The covariance matrix for the REML estimates of fixed effects is:

```r
vcov(lmm1)
```

```
##             (Intercept)       genderM   attitudepol
## (Intercept)   229.67362 -2.195819e+02 -2.018345e+01
## genderM      -219.58189  4.391638e+02  6.451438e-15
## attitudepol   -20.18345  6.451438e-15  4.036690e+01
```

BLUPs for subject-specific intercepts:

```r
random.effects(lmm1)
```

```
##    (Intercept)
## F1  -13.575831
## F2   10.170522
## F3    3.405309
## M3   27.960288
## M4    4.739325
## M7  -32.699613
```

Residuals:

```r
data$frequency-fitted(lmm1)
```

```
##          F1          F1          F1          F1          F1          F1
## -10.1086926 -38.9110735  61.6913074  16.2889265 -19.5086926  43.4889265
##          F1          F1          F1          F1          F1          F1
##  27.3913074  33.3889265   8.4913074   8.9889265 -42.2086926 -12.7110735
##          F1          F1          F3          F3          F3          F3
## -26.9110735 -68.6086926 -10.6898326 -23.0922136  -3.5898326  -9.3922136
##          F3          F3          F3          F3          F3          F3
##  26.6101674   5.6077864  35.0101674  46.4077864  -7.7898326  -7.8922136
##          F3          F3          F3          F3          M4          M4
## -13.8898326  18.4077864   4.0077864 -54.8898326 -22.2262298 -29.3286108
##          M4          M4          M4          M4          M4          M4
##  96.0737702 -38.0286108 -20.7262298  60.6713892  60.4737702   9.9713892
##          M4          M4          M4          M4          M4          M4
## -31.1262298 -26.0286108 -22.9262298 -16.7286108  -6.9286108  -6.4262298
##          M7          M7          M7          M7          M7          M7
##  -9.3872916 -16.3896725 -13.2872916 -11.1896725  -9.5872916  -5.2896725
##          M7          M7          M7          M7          M7          M7
##   1.6127084   4.5103275  -1.7872916 -12.5896725  13.3127084  -7.2896725
##          M7          M7          F2          F2          F2          F2
##   8.9103275  12.1127084 -14.4550462 -35.8574271  -0.8550462  -7.4574271
##          F2          F2          F2          F2          F2          F2
##  42.2449538  34.6425729  -3.9550462  29.0425729  30.5449538  27.0425729
##          F2          F2          F2          F2          M3          M3
## -39.1550462 -41.2574271  13.8425729 -19.9550462  -2.3471929  12.6504261
##          M3          M3          M3          M3          M3          M3
## -13.7471929  23.5504261   4.0528071   9.9504261  51.3528071  14.7504261
##          M3          M3          M3          M3          M3          M3
##   4.5528071 -19.6495739  -9.4471929 -18.1495739 -15.0495739  -2.8471929
## attr(,"label")
## [1] "Fitted values"
```

Question 3

Fit model 2.

```r
lmm2 <- lme(frequency ~ gender + attitude + gender*attitude, random = ~1 | subject,
            data = data, method='REML')
summary(lmm2)
```

```
## Linear mixed-effects model fit by REML
##  Data: data
##        AIC     BIC    logLik
##   799.8018 814.094 -393.9009
##
## Random effects:
##  Formula: ~1 | subject
##         (Intercept) Residual
## StdDev:    24.46382 29.04716
##
## Fixed effects: frequency ~ gender + attitude + gender * attitude
##                          Value Std.Error DF   t-value p-value
## (Intercept)          260.68571 15.481307 76 16.838740  0.0000
## genderM             -116.19524 21.893875  4 -5.307203  0.0061
## attitudepol          -27.40000  8.964149 76 -3.056620  0.0031
## genderM:attitudepol   14.79524 12.677221 76  1.167073  0.2468
##  Correlation:
##                     (Intr) gendrM atttdp
## genderM             -0.707
## attitudepol         -0.290  0.205
## genderM:attitudepol  0.205 -0.290 -0.707
##
## Standardized Within-Group Residuals:
##        Min         Q1        Med         Q3        Max
## -2.2344163 -0.5454437 -0.1646159  0.4697182  3.1800944
##
## Number of Observations: 84
## Number of Groups: 6
```

Likely ration test.

$H_0 : \beta_{gender*attitude} = 0$, $H_1 : \beta_{gender*attitude} \neq 0$

```r
lmm.1 = lme(frequency ~ gender + attitude ,
              random = ~ 1|subject, data = data, method = "ML")
lmm.2 = lme(frequency ~ gender + attitude + gender*attitude, random = ~1 | subject,
            data = data, method='ML')
anova(lmm.2, lmm.1)
```

```
##       Model df      AIC      BIC    logLik   Test  L.Ratio p-value
## lmm.2     1  6 826.2508 840.8357 -407.1254
## lmm.1     2  5 825.6363 837.7904 -407.8182 1 vs 2 1.385523  0.2392
```

As p-value equals to 0.2392 which is greater than 0.05, we fail to reject the null hypothesis and conclude that $\beta_{gender*attitude} = 0$, thus smaller model is better.

Question 4

Fit the model 3.

```r
lmm3 <- lme(frequency ~ gender + attitude, random = ~1 + attitude| subject,
            data = data, method='REML')
summary(lmm3)
```

```
## Linear mixed-effects model fit by REML
##  Data: data
##        AIC      BIC    logLik
##   810.0805 826.8416 -398.0402
##
## Random effects:
##  Formula: ~1 + attitude | subject
##  Structure: General positive-definite, Log-Cholesky parametrization
##             StdDev       Corr
## (Intercept) 24.458032213 (Intr)
## attitudepol  0.003285569 0
## Residual    29.115372269
##
## Fixed effects: frequency ~ gender + attitude
##                  Value Std.Error DF   t-value p-value
## (Intercept)  256.98691 15.154987 77 16.957250  0.0000
## genderM     -108.79762 20.956235  4 -5.191659  0.0066
## attitudepol  -20.00238  6.353495 77 -3.148248  0.0023
##  Correlation:
##             (Intr) gendrM
## genderM     -0.691
## attitudepol -0.210  0.000
##
## Standardized Within-Group Residuals:
##        Min         Q1        Med         Q3        Max
## -2.3564422 -0.5658319 -0.2011979  0.4617896  3.2997610
##
## Number of Observations: 84
## Number of Groups: 6
```

As

$$Y_{ij} = \beta_0 + X_{ij}^T\beta + b_{1i} + b_{2i}t_{ij} + \epsilon_{ij}$$

where $b_{1i} \sim N(0, g_{11})$, $b_{2i} \sim N(0, g_{22})$ are random intercept and random slope with $cov(b_{1i}, b_{2i}) = g_{12}$ and $\epsilon_{ij} \sim N(0, \sigma^2)$.

So

$$var(Y_i) = g_{11} + 2t_{ij} + t_{ij}^2 g_{22} + \sigma^2$$

when the attitude is `inf`, $t_{ij} = 0$, so

$$var(Y_i) = g_{11} + \sigma^2$$

when the attitude is `pol`, $t_{ij} = 1$, so

$$var(Y_i) = g_{11} + 2g_{12} + g_{22} + \sigma^2$$

And

$$cov(Y_{ij}, Y_{ik}) = g_{11} + (t_{ij} + t_{ik})g_{12} + t_{ij}t_{ik}g_{22}$$

when attitude of two observations are both `inf`, $t_{ij} = t_{ik} = 0$, so

$$cov(Y_{ij}, Y_{ik}) = g_{11}$$

when attitude of two observations are both `pol`, $t_{ij} = t_{ik} = 1$, so

$$cov(Y_{ij}, Y_{ik}) = g_{11} + 2g_{12} + g_{22}$$

when attitude of two observations are `pol` and `inf`, $t_{ij} = 0, t_{ik} = 1$, so

$$cov(Y_{ij}, Y_{ik}) = g_{11} + g_{12}$$

```
#VarCorr(lmm3)
g11 = as.numeric(VarCorr(lmm3)[1,2])^2
g22 = as.numeric(VarCorr(lmm3)[2,2])^2
g12 = as.numeric(VarCorr(lmm3)[2,3])
sigma = as.numeric(VarCorr(lmm3)[3,2])
g11 + sigma^2
```

```
## [1] 1445.9
```

```
g11 + 2*g12 + g22 + sigma^2
```

```
## [1] 1445.9
```

```
g11
```

```
## [1] 598.1953
```

```
g11 + 2*g12 + g22
```

```
## [1] 598.1954
```

```
g11 + g12
```

```
## [1] 598.1953
```

For the same attitude and the attitude is `inf`:

$$A = \begin{bmatrix} g_{11} + \sigma^2 & g_{11} & \cdots & g_{11} \\ g_{11} & g_{11} + \sigma^2 & \cdots & g_{11} \\ \vdots & \vdots & \ddots & \vdots \\ g_{11} & g_{11} & \cdots & g_{11} + \sigma^2 \end{bmatrix}$$

For the same attitude and the attitude is `pol`:

$$B = \begin{bmatrix} g_{11} + 2g_{12} + g_{22} + \sigma^2 & g_{11} + 2g_{12} + g_{22} & \cdots & g_{11} + 2g_{12} + g_{22} \\ g_{11} + 2g_{12} + g_{22} & g_{11} + 2g_{12} + g_{22} + \sigma^2 & \cdots & g_{11} + 2g_{12} + g_{22} \\ \vdots & \vdots & \ddots & \vdots \\ g_{11} + 2g_{12} + g_{22} & g_{11} + 2g_{12} + g_{22} & \cdots & g_{11} + 2g_{12} + g_{22} + \sigma^2 \end{bmatrix}$$

For different attitudes:

$$C = \begin{bmatrix} g_{11} + g_{12} & g_{11} + g_{12} & \cdots & g_{11} + g_{12} \\ g_{11} + g_{12} & g_{11} + g_{12} & \cdots & g_{11} + g_{12} \\ \vdots & \vdots & \ddots & \vdots \\ g_{11} + g_{12} & g_{11} + g_{12} & \cdots & g_{11} + g_{12} \end{bmatrix}$$

The covariance matrix for a subject Y_i is:

$$\begin{bmatrix} A & C \\ C & B \end{bmatrix}$$

$$= \begin{bmatrix} g_{11}+\sigma^2 & g_{11} & \dots & g_{11} & g_{11}+g_{12} & g_{11}+g_{12} & \dots & g_{11}+g_{12} \\ g_{11} & g_{11}+\sigma^2 & \dots & g_{11} & g_{11}+g_{12} & g_{11}+g_{12} & \dots & g_{11}+g_{12} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ g_{11} & g_{11} & \dots & g_{11}+\sigma^2 & g_{11}+g_{12} & g_{11}+g_{12} & \dots & g_{11}+g_{12} \\ g_{11}+g_{12} & g_{11}+g_{12} & \dots & g_{11}+g_{12} & g_{11}+2g_{12}+g_{22}+\sigma^2 & g_{11}+2g_{12}+g_{22} & \dots & g_{11}+2g_{12}+g_{22} \\ g_{11}+g_{12} & g_{11}+g_{12} & \dots & g_{11}+g_{12} & g_{11}+2g_{12}+g_{22} & g_{11}+2g_{12}+g_{22}+\sigma^2 & \dots & g_{11}+2g_{12}+g_{22} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ g_{11}+g_{12} & g_{11}+g_{12} & \dots & g_{11}+g_{12} & g_{11}+2g_{12}+g_{22} & g_{11}+2g_{12}+g_{22} & \dots & g_{11}+2g_{12}+g_{22}+\sigma^2 \end{bmatrix}$$

$$= \begin{bmatrix} 1445.92 & 598.1953 & \dots & 598.1953 & 598.1953 & 598.1953 & \dots & 598.1953 \\ 598.1953 & 1445.92 & \dots & 598.1953 & 598.1953 & 598.1953 & \dots & 598.1953 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 598.1953 & 598.1953 & \dots & 1445.92 & 598.1953 & 598.1953 & \dots & 598.1953 \\ 598.1953 & 598.1953 & \dots & 598.1953 & 1445.9 & 598.1954 & \dots & 598.1954 \\ 598.1953 & 598.1953 & \dots & 598.1953 & 598.1954 & 1445.9 & \dots & 598.1954 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 598.1953 & 598.1953 & \dots & 598.1953 & 598.1954 & 598.1954 & \dots & 1445.9 \end{bmatrix}$$

As 598.1953 and 598.1954 are very closed, so the covariance structure for subject Y_i can be approximate to compound symmetry.

The fix effect, random effect and BLUP:

```r
fixed.effects(lmm3)
```

```
## (Intercept)     genderM attitudepol
##   256.98691  -108.79762   -20.00238
```

```r
random.effects(lmm3)
```

```
##    (Intercept)   attitudepol
## F1  -13.575831 -8.408891e-07
## F2   10.170522  1.499413e-07
## F3    3.405308 -2.981919e-07
## M3   27.960288  1.009764e-06
## M4    4.739325  7.794162e-07
## M7  -32.699612 -8.000404e-07
```

So the fixed effect for the first female subject in scenario 1 with polite attitude is 256.987-20.002 = 236.985 and the random effect for the first female subject in scenario 1 with polite attitude is -13.575831-(8.408891e-07) = -13.576. And the BLUP for the intercept is -13.575831, the BLUP for the slop is 8.408891e-07. Corresponding $\hat{Y}_i$ equals to 236.985-13.576 = 223.409.