p8131_hw7_xy2395

Jack Yan

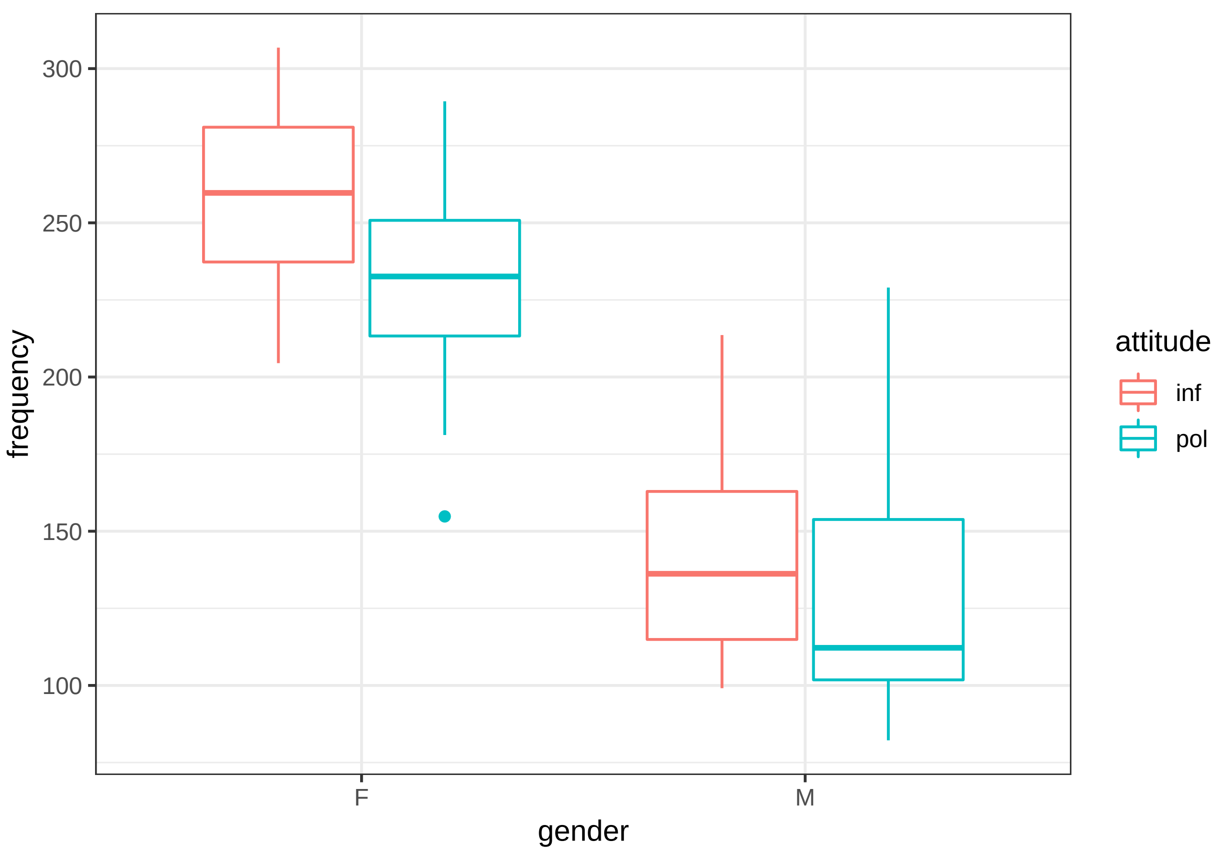
4/11/2019

```r
# Import data
polite_df <-
  read_csv('../hw7/HW7-politeness_data.csv') %>%
  as.tibble() %>% janitor::clean_names()
```

1. Exploratory Analysis

Provide boxplots to show the relation between gender/attitude and pitch.

```r
# boxplots
polite_df %>%
  ggplot(aes(x = gender, y = frequency, color = attitude)) +
    geom_boxplot() +
    theme_bw()
```



Males generally tend to have lower pitch than females. Within each gender, informal attitude (inf) tends to have higher pitch than formal attitude (pol).

2. Mixed Effects Model with Random Intercept

Fit a mixed effects model with random intercepts for different subjects (gender and attitude being the fixed effects).

```
lmm = lme(frequency ~ gender + attitude, random = ~1 | subject, method = "REML", data = polite_df)

VarCorr(lmm)
```

```
## subject = pdLogChol(1)
##             Variance StdDev
## (Intercept) 598.1953 24.45803
## Residual    847.7049 29.11537
```

```
# var(Yi) = 598.1953 + 847.7049 = 1445.9
var = as.numeric(VarCorr(lmm)[[1]]) + as.numeric(VarCorr(lmm)[[2]]); var
```

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

```
# cov(Yij,Yik) = 598.2
cov = as.numeric(VarCorr(lmm)[[1]]); cov
```

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

The covariance matrix for a subject Y_i follows a compound symmetry pattern with $var(Y_{ij}) = 1445.9$ and $cov(Y_{ij}, Y_{ik}) = 598.2$. There are 14 measurements within each subject, so the covariance matrix is a 14*14 matrix with $var(Y_{ij}) = 1445.9$ as diagonal values and $cov(Y_{ij}, Y_{ik}) = 598.2$ as off-diagonal values.

$$var(Y_i) = \begin{bmatrix} 1445.9 & 598.2 & \dots & 598.2 \\ 598.2 & 1445.9 & & \\ \vdots & & \ddots & \vdots \\ 598.2 & & \dots & 1445.9 \end{bmatrix}_{14\times14}$$

```
# covariance matrix for the REML estimates of fixed effects
vcov(lmm)
```

```
##             (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
```

The covariance matrix for the REML estimates of fixed effects is shown above.

```
# BLUPs for subject-specific intercepts
random.effects(lmm)
```

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

The BLUPs for subject-specific intercepts are shown above.

```
resid_lmm = polite_df$frequency - fitted(lmm)
resid_lmm
```

```
##          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"
```

Residuals are the deviations from subject-specific mean. The residuals are shown above.

3. Likelihood Ratio Test for the Interaction Term

Fit a mixed effects model with intercepts for different subjects (gender, attitude and their interaction being the fixed effects).

```r
lmm_small = lme(frequency ~ gender + attitude, random = ~1 | subject, method = "ML", data = polite_df)
lmm_large = lme(frequency ~ gender * attitude, random = ~1 | subject, method = "ML", data = polite_df)

anova(lmm_small, lmm_large)
```

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

We use maximum likelihood method to fit the model. The p-value for Likelihood ratio test is $0.2392 > 0.05$, so we fail to reject the null and conclude that the interaction term is not significantly associated with pitch, at the significance level of 0.05.

4. Mixed Effects Model with Random Intercept and Slope

Model: $Y_{ij} = \beta_1 + \beta_2 * I(gender = Female) + \beta_3 * t_{ij} + b_{1i} + b_{2i} * t_{ij} + \epsilon_{ij}$, where $t_{ij} = I(attitude_{ij} = pol)$, β_1 is the fixed intercept, β_2 and β_3 are slopes, and b_{1i} and b_{2i} are random intercept and random slope, respectively.

```r
lmm2 = lme(frequency ~ gender + attitude, random = ~1 + attitude | subject, method = "REML", data = poli

VarCorr(lmm2)
```

```
## subject = pdLogChol(1 + attitude)
##             Variance     StdDev      Corr
## (Intercept) 5.981953e+02 24.458032213 (Intr)
## attitudepol 1.079496e-05  0.003285569 0
## Residual    8.477049e+02 29.115372269
```

```r
getVarCov(lmm2)
```

```
## Random effects variance covariance matrix
##             (Intercept) attitudepol
## (Intercept)  5.9820e+02  1.3002e-05
## attitudepol  1.3002e-05  1.0795e-05
##   Standard Deviations: 24.458 0.0032856
```

We get $g_{11} = 598.2$, $g_{22} = 1.079496 * 10^{-5}$, $g_{12} = 0$, and $\sigma^2 = 847.7$.

Variance: $var(Y_{ij}) = g_{11} + 2t_{ij}g_{12} + t_{ij}^2 g_{22} + \sigma^2$

1. If $t_{ij} = 1$, $var(Y_{ij}) = g_{11} + 2t_{ij}g_{12} + t_{ij}^2 g_{22} + \sigma^2 = 598.2 + 2 * 0 + 1.079496 * 10^{-5} + 847.7 = 1445.9$
2. If $t_{ij} = 0$, $var(Y_{ij}) = g_{11} + \sigma^2 = 598.2 + 847.7 = 1445.9$

The general formula of covariance between Y_{ij} and Y_{ik} is

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

1. If $t_{ij} = 0$ and $t_{ik} = 0$, $cov(Y_{ij}, Y_{ik}) = g_{11} = 598.2$.
2. If $t_{ij} = 0$ and $t_{ik} = 1$, $cov(Y_{ij}, Y_{ik}) = g_{11} + g_{12} = 598.2 + 0 = 598.2$.
3. If $t_{ij} = 1$ and $t_{ik} = 0$, $cov(Y_{ij}, Y_{ik}) = g_{11} + g_{12} = 598.2 + 1.079496 * 10^{-5} = 598.2$.
4. If $t_{ij} = 1$ and $t_{ik} = 1$, $cov(Y_{ij}, Y_{ik}) = g_{11} + 2g_{12} + g_{22} = 598.2 + 2 * 0 + 1.079496 * 10^{-5} = 598.2$.

So we get $var(Y_{ij}) = 1445.9$ and $cov(Y_{ij}, Y_{ik}) = 598.2$ for any different j and k. **Matrix Notation:**

Suppose there are 14 observations within an individual, and for observations 1-7 $t_{ij} = 0$, and for observations 8-14 $t_{ij} = 1$. The covariance matrix is as follows:

$$A = \begin{bmatrix} g_{11} + \sigma^2 & g_{11} & \cdots & g_{11} \\ g_{11} & g_{11} + \sigma^2 & & \vdots \\ \vdots & & \ddots & g_{11} \\ g_{11} & \cdots & g_{11} & g_{11} + \sigma^2 \end{bmatrix}_{7\times 7},$$

$$B = \begin{bmatrix} g_{11} + 2t_{ij}g_{12} + t_{ij}^2 g_{22} + \sigma^2 & g_{11} + 2g_{12} + g_{22} & \cdots & g_{11} + 2g_{12} + g_{22} \\ g_{11} + 2g_{12} + g_{22} & g_{11} + 2t_{ij}g_{12} + t_{ij}^2 g_{22} + \sigma^2 & & \vdots \\ \vdots & & \ddots & g_{11} + 2g_{12} + g_{22} \\ g_{11} + 2g_{12} + g_{22} & \cdots & g_{11} + 2g_{12} + g_{22} & g_{11} + 2t_{ij}g_{12} + t_{ij}^2 g_{22} + \sigma^2 \end{bmatrix}_{7\times 7},$$

$$C = \begin{bmatrix} g_{11} + g_{12} & g_{11} + g_{12} & \cdots & g_{11} + g_{12} \\ g_{11} + g_{12} & g_{11} + g_{12} & & \vdots \\ \vdots & & \ddots & g_{11} + g_{12} \\ g_{11} + g_{12} & \cdots & g_{11} + g_{12} & g_{11} + g_{12} \end{bmatrix}_{7\times 7}$$

$$var(Y_i) = \begin{bmatrix} A & C \\ C & B \end{bmatrix}_{14\times 14} = \begin{bmatrix} 1445.9 & 598.2 & \dots & 598.2 \\ 598.2 & 1445.9 & & \\ \vdots & & \ddots & \vdots \\ 598.2 & & \dots & 1445.9 \end{bmatrix}_{14\times 14}$$

The random slope has small variance, so it captures little variation in the data. So this model does not do a much better job than a random intercept model. The covariance structure for this model is thus numerically a compound symmetry structure as in Question 2.

```
fixed.effects(lmm2)
```

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

```
fixed_effect = 256.98691 + -20.00238; fixed_effect
```

```
## [1] 236.9845
```

```
random.effects(lmm2)
```

```
##    (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
```

```
random_effect = -13.575831 + -8.408891 * 10^(-07); random_effect
```

```
## [1] -13.57583
```

```
BLUP = fixed_effect + random_effect; BLUP
```

```
## [1] 223.4087
```

The fixed effect for intercept is 256.98691, and for attitude being polite is -20.00238; so the fixed effects in total is 236.9845. The random effect for intercept is -13.575831 and for attitude being polite is $-8.408891 \times 10^{-07}$; so the random effects in total is -13.57583. The BLUP is the summation of fixed effects and random effects, 223.4087.

5. ?

```
library(lme4)
```

```
## Loading required package: Matrix

##
## Attaching package: 'Matrix'

## The following object is masked from 'package:tidyr':
##
##     expand

##
## Attaching package: 'lme4'

## The following object is masked from 'package:nlme':
##
##     lmList
```

```r
lmm3 = lmer(frequency ~ gender + attitude + (1 | subject) +  (1 | scenario), data = polite_df)
summary(lmm3)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: frequency ~ gender + attitude + (1 | subject) + (1 | scenario)
##    Data: polite_df
##
## REML criterion at convergence: 784.1
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.2690 -0.6331 -0.0878  0.5204  3.5326
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  scenario (Intercept) 224.5    14.98
##  subject  (Intercept) 613.2    24.76
##  Residual             637.8    25.25
## Number of obs: 84, groups:  scenario, 7; subject, 6
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  256.987     16.101  15.961
## genderM     -108.798     20.956  -5.192
## attitudepol  -20.002      5.511  -3.630
##
## Correlation of Fixed Effects:
##             (Intr) gendrM
## genderM     -0.651
## attitudepol -0.171  0.000
```

```r
# Variance of Y_ij
224.5 + 613.2 + 637.8
```

```
## [1] 1475.5
```

```r
# Covariance of Y_ij and Y_ik with same scenario
224.5 + 613.2
```

```
## [1] 837.7
```

```r
# Covariance of Y_ij and Y_ik with different scenarios
613.2
```

```
## [1] 613.2
```

Denote variance of subject specific intercept as σ_1^2, variance of scenario specific intercept as σ_2^2, and variance of residuals as σ^2. We have $\sigma_1^2 = 613.2$, $\sigma_2^2 = 224.5$, and $\sigma^2 = 637.8$.

$A = \begin{bmatrix} \sigma_1^2 + \sigma_2^2 + \sigma^2 & \sigma_1^2 + \sigma_2^2 \\ \sigma_1^2 + \sigma_2^2 & \sigma_1^2 + \sigma_2^2 + \sigma^2 \end{bmatrix} = \begin{bmatrix} 1475.5 & 837.7 \\ 837.7 & 1475.5 \end{bmatrix}$ is the variance-covariance matrix of 2 observations with the same scenario within an individual.

$B = \begin{bmatrix} \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 \end{bmatrix} = \begin{bmatrix} 613.2 & 613.2 \\ 613.2 & 613.2 \end{bmatrix}$ is the covariance matrix of observations with different scenarios within an individual.

We get $var(Y_i) = \begin{bmatrix} A & B & \dots & B \\ B & A & & \vdots \\ \vdots & & \ddots & B \\ B & \dots & B & A \end{bmatrix}_{14\times 14}$.

Interpretation:

The mean pitch will on avarage be 20.002 units lower for polite attitude than informal attitude over all observations, within the same gender.