P8131_hw7

Shan Jiang

(1) Exploratory analysis

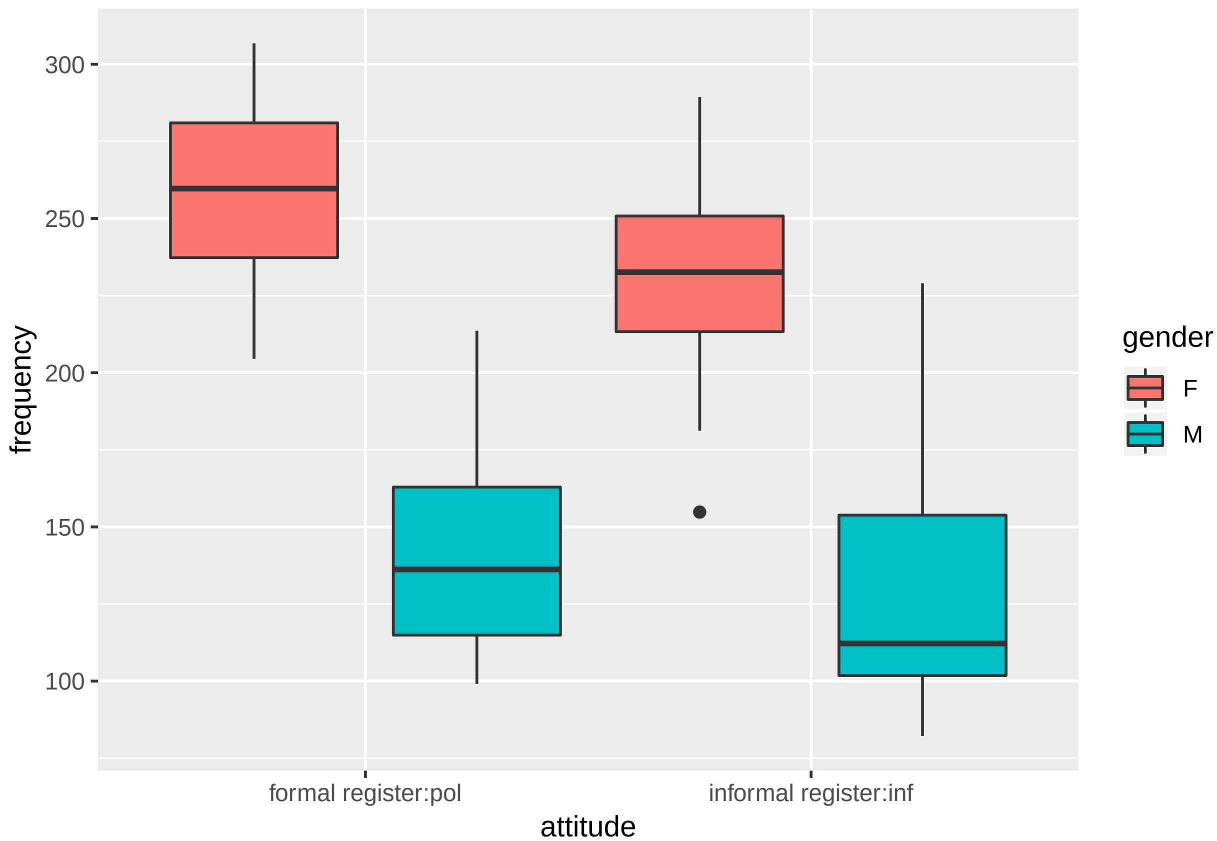
Import data

```
library(tidyverse)
library(nlme)
pol_df = read.csv("./HW7-politeness_data.csv")
pol_df$subject <- as.factor(pol_df$subject)
head(pol_df)
```

```
##   subject gender scenario attitude frequency
## 1      F1      F        1      pol     213.3
## 2      F1      F        1      inf     204.5
## 3      F1      F        2      pol     285.1
## 4      F1      F        2      inf     259.7
## 5      F1      F        3      pol     203.9
## 6      F1      F        3      inf     286.9
```

Boxplots to show the relationship

```
ggplot(pol_df, aes(x = attitude, y = frequency, fill = gender)) +
  geom_boxplot(outlier.shape= 16,
               outlier.size=2, position = position_dodge(1)) +
  scale_x_discrete(labels = c("formal register:pol","informal register:inf"))
```



From this plot, attitude inf category has an overall higher value than pol group, meantime, we can find that there is significant difference between male and female groups, female also has presented a higehr frequency in both two attitude categories.

(2) Mixed eff ects model with random interceptsfor di erent subjects (gender and attitude bets).

```
## Grouped data: 14 for each subject
pol_df %>%
  group_by( subject) %>%
  summarise(n = n())
```

```
## # A tibble: 6 x 2
##   subject     n
##   <fct>   <int>
## 1 F1         14
## 2 F2         14
## 3 F3         14
## 4 M3         14
## 5 M4         14
## 6 M7         14
```

Since the pitch measurements are typically correlated for same subject and in the same scenario, we have our grouping variable as `subject` and `attitude`.

Mixed Effects Model:

$$\overbrace{Y_i}^{84\times 1} = \overbrace{\underbrace{X_i}_{84\times 2}\underbrace{\beta}_{2\times 1}}^{84\times 1} + \overbrace{\underbrace{Z_i}_{84\times 14}\underbrace{b_i}_{14\times 1}}^{84\times 1} + \overbrace{\epsilon_i}^{84\times 1}, \text{i = 1, 2, ... , m}$$

where:

- Y_i is 84×1 vector;
- X is 84×2 matrix of the 2 predictor variables: `attitude` and `gender`;
- β is a $2\times$ 1column vector of the fixed-effects regression coefficients;
- Z_i is the 84×12 design matrix for random effects;
- b_i is the 12×1 random effects;
- e_i is error vector.

Distributional assumptions: bi and i are independent

$$b_i \sim N(0, G), e_i \sim N(0, \sigma^2 I)$$

Model 1: fit a random intercept model

Random intercepts, with linear attitude and gender effect

```
# fit a random intercept model
LMM1 <- lme(frequency ~  gender + attitude,
             random = ~ 1 | subject,  data = pol_df, method = 'REML')
summary(LMM1) # pay attention to: random effects, fixed effects,
```

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

- The fixed-effect intercept coefficient for dummy variable labelled `attitudepol` is $\hat{\beta}_0$ = -20.00238, represents an estimate of an average level of pitch frequency, which are the baseline category for female individuals in informal politeness group.

- ```
  # 2.1 The covariance matrix for subject Y_i:
  LMM1$sigma
  VarCorr(LMM1) # covariance estimates for random effects and variance for residuals
  sigma_b = as.numeric(VarCorr(LMM1)[1,2])^2
  Varyij = as.numeric(VarCorr(LMM1)[2,2])^2 + sigma_b  # variance of yij
  Varyij
  ```

- The second panel of summary displays estimates of the variance and covariance parameters for the random effects, in the form of standard deviations and correlations, the term labelled Residual is the estimate of

$$Var(Y_i) = 847.7049 + 598.1953 = 1445.9$$

- Because we only have one random intercept, $u \sim N(0, G)$, by assuming that the random effects are independent,

- Thus, the estimated variance-covariance matrix of the random effects is therefore:

$$G = \begin{pmatrix} 1445.9 & 598.1953 & \cdots & 598.1953 \\ 598.1953 & 1445.9 & \cdots & 598.1953 \\ \vdots & \vdots & \ddots & \vdots \\ 598.1953 & 1445.9 & \cdots & 598.1953 \end{pmatrix}_{14\times 14}$$

```r
# 2.2 The covariance matrix for REML fixed effects estimates (inv fisher info)
vcov(LMM1) # covariance for fixed effects estimates (inv fisher info)
```

```
##             (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 REML fixed effects estimates:

$$R = \begin{pmatrix} 229.67362 & -219.5819 & -20.18345 \\ -219.5819 & 439.1638 & 6.451438\times 10^{-15} \\ -20.18345 & 6.451438\times 10^{-15} & 40.369 \end{pmatrix}$$

```r
# 2.3 BLUPs for subject-specific intercepts
random.effects(LMM1) # ordered random effects, BLUP (in this case, just b_i)
```

```
##    (Intercept)
## F1  -13.575831
```
```

```
## F2   10.170522
## F3    3.405309
## M3   27.960288
## M4    4.739325
## M7  -32.699613
```

The Best Linear Unbiased Predictor(BLUP) should be above.

```r
# 2.4 Residuals
pol_df$frequency-fitted(LMM1) # residuals
```

```
##          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"
```

(3) Fit with additional interaction term

Fit a model with interaction term attitude * gender.

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

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

The Likelihood Ratio Test (LRT) of fixed effects requires the models be fit with by MLE (use REML=FALSE for linear mixed models.)

```r
LMM.1 <- lme(frequency ~ attitude + gender,
             random = ~ 1 | subject,  data = pol_df,  method = 'ML')
summary(LMM.1)
```

```
## Linear mixed-effects model fit by maximum likelihood
##  Data: pol_df
##        AIC      BIC    logLik
##   825.6363 837.7904 -407.8182
##
## Random effects:
##  Formula: ~1 | subject
##         (Intercept) Residual
## StdDev:    19.47793 28.92813
##
## Fixed effects: frequency ~ attitude + gender
##                  Value Std.Error DF   t-value p-value
## (Intercept)  256.98690 12.733461 77 20.182016  0.0000
## attitudepol  -20.00238  6.428474 77 -3.111529  0.0026
## genderM     -108.79762 17.424678  4 -6.243881  0.0034
##  Correlation:
##             (Intr) atttdp
## attitudepol -0.252
## genderM     -0.684  0.000
##
## Standardized Within-Group Residuals:
##        Min         Q1        Med         Q3        Max
## -2.3945323 -0.5728233 -0.2228552  0.4871624  3.3290916
##
## Number of Observations: 84
## Number of Groups: 6
```

```r
LMM.2 <- lme(frequency ~ attitude + gender + attitude * gender,
             random = ~ 1 | subject,  data = pol_df,  method = 'ML')
summary(LMM.2)
```

```
## Linear mixed-effects model fit by maximum likelihood
##  Data: pol_df
##        AIC      BIC    logLik
##   826.2508 840.8357 -407.1254
##
```

```
## Random effects:
##  Formula: ~1 | subject
##         (Intercept) Residual
## StdDev:    19.50493 28.67234
##
## Fixed effects: frequency ~ attitude + gender + attitude * gender
##                         Value Std.Error DF   t-value p-value
## (Intercept)          260.68571 13.200754 76 19.747790  0.0000
## attitudepol          -27.40000  9.066991 76 -3.021951  0.0034
## genderM             -116.19524 18.668685  4 -6.224072  0.0034
## attitudepol:genderM   14.79524 12.822662 76  1.153835  0.2522
##  Correlation:
##                     (Intr) atttdp gendrM
## attitudepol         -0.343
## genderM             -0.707  0.243
## attitudepol:genderM  0.243 -0.707 -0.343
##
## Standardized Within-Group Residuals:
##        Min         Q1        Med         Q3        Max
## -2.2856421 -0.5245601 -0.1718554  0.4929026  3.2293520
##
## Number of Observations: 84
## Number of Groups: 6
```

Compare Two models using LR test

```r
anova (LMM.1, LMM.2)
```

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

The 1 vs 2 Likelihood Ratio test results proved that the the new model with interaction term for `attitude` and `gender` is not significantly associated with pitch since P-value 0.2392 is larger than the common cut off alpha level of .05.

(4) Model 3: Random Intercepts and random slope

```r
# fit a random intercept and slope model
LMM3 <- lme(frequency ~ attitude + gender, random = ~ 1 + attitude| subject, data = pol_df, method = 'REML')
summary(LMM3)
```

```
## Linear mixed-effects model fit by REML
##  Data: pol_df
##        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  0.003287353 0
## Residual    29.115372269
##
## Fixed effects: frequency ~ attitude + gender
##                 Value Std.Error DF   t-value p-value
## (Intercept)  256.98691 15.154987 77 16.957250  0.0000
## attitudepol  -20.00238  6.353495 77 -3.148248  0.0023
```

```
## genderM     -108.79762 20.956235  4 -5.191659  0.0066
##  Correlation:
##             (Intr) atttdp
## attitudepol -0.210
## genderM     -0.691  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
```

4.1 Covariance structure for a subject Y_i:

```
# Covariance structure for a subject
VarCorr(LMM3) # covariance estimates for random effects and variance for residuals
```

```
## subject = pdLogChol(1 + attitude)
##             Variance     StdDev       Corr
## (Intercept) 5.981953e+02 24.458032213 (Intr)
## attitudepol 1.080669e-05  0.003287353 0
## Residual    8.477049e+02 29.115372269
```

```
LMM3$sigma # std for residuals
```

```
## [1] 29.11537
```

```
vcov(LMM3)
```

```
##             (Intercept)   attitudepol       genderM
## (Intercept)   229.67362 -2.018345e+01 -2.195819e+02
## attitudepol   -20.18345  4.036690e+01  1.048509e-14
## genderM      -219.58190  1.048509e-14  4.391638e+02
```

```
#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])
hat_sigma = as.numeric(VarCorr(LMM3)[3,2])
```

```
# (1) when attitude of two observations are both inf
g11 + (hat_sigma)^2 # var
```

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

```
g11 # cov
```

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

```
# (2) when attitude of two observations are both pol
g11 + 2*g12 + g22 + (hat_sigma)^2  # var
```

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

```
g11 + 2*g12 +g22 #cov
```

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

```
# (3) when attitude of two observations are pol and inf
g11 + g12 # cov
```

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

- The second part of summary panel displays estimates of the variance and covariance parameters for the random effects Y_i, in the form of standard deviations and correlations.
- As 598.1953 and 598.1954 are very closed, so the covariance structure for subject Yi can be approximate to compound symmetry.
- In total, there are 3 cases:

1. For the same attitude and the attitude is inf:

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

2. For the same attitude and the attitude is pol:

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

3. For different attitudes:

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

- Thus, The covariance matrix for a subject Yi is

$$Vcov(Y_{ij}) = \begin{pmatrix} A & C \\ C & B \end{pmatrix} = \begin{pmatrix} 1445.92 & 598.1953 & \cdots & 598.1953 & 598.1953 & 598.1953 & \cdots & 598.1953 \\ 598.1953 & 1445.92 & \cdots & 598.1953 & 598.1953 & 598.1953 & \cdots & 598.1953 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 598.1953 & 598.1953 & 598.1953 & 1445.92 & 598.1953 & 598.1953 & 598.1953 & 598.1953 \\ 598.1953 & 598.1953 & 598.1953 & 598.1953 & 1445.92 & 598.1953 & 598.1953 & 598.1953 \\ 598.1953 & 598.1953 & 598.1953 & 598.1953 & 598.1953 & 1445.92 & 598.1953 & 598.1953 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 598.1953 & 598.1953 & \cdots & 598.1953 & 598.1953 & 598.1953 & \cdots & 1445.92 \end{pmatrix}$$

4.2 Fixed Effect

```
fixed.effects(LMM3) # fixed effects coeff
```

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

4.3 Random Effect

```
random.effects(LMM3)
```

```
##    (Intercept)   attitudepol
## F1  -13.575831 -8.418005e-07
## F2   10.170522  1.501026e-07
## F3    3.405308 -2.985164e-07
## M3   27.960288  1.010857e-06
## M4    4.739325  7.802621e-07
## M7  -32.699612 -8.009045e-07
```

4.4 The Blup for first female in S1 with polite attitude.

```r
random.effects(LMM3) # ordered random effects, BLUP (in this case, just b_i)
```

```
##    (Intercept)   attitudepol
## F1  -13.575831 -8.418005e-07
## F2   10.170522  1.501026e-07
## F3    3.405308 -2.985164e-07
## M3   27.960288  1.010857e-06
## M4    4.739325  7.802621e-07
## M7  -32.699612 -8.009045e-07
```

```r
fitted(LMM3)
```

```
##        F1        F1        F1        F1        F1        F1        F1
## 223.40869 243.41107 223.40869 243.41107 223.40869 243.41107 223.40869
##        F1        F1        F1        F1        F1        F1        F1
## 243.41107 223.40869 243.41107 223.40869 243.41107 243.41107 223.40869
##        F3        F3        F3        F3        F3        F3        F3
## 240.38983 260.39221 240.38983 260.39221 240.38983 260.39221 240.38983
##        F3        F3        F3        F3        F3        F3        F3
## 260.39221 240.38983 260.39221 240.38983 260.39221 260.39221 240.38983
##        M4        M4        M4        M4        M4        M4        M4
## 132.92623 152.92861 132.92623 152.92861 132.92623 152.92861 132.92623
##        M4        M4        M4        M4        M4        M4        M4
## 152.92861 132.92623 152.92861 132.92623 152.92861 152.92861 132.92623
##        M7        M7        M7        M7        M7        M7        M7
##  95.48729 115.48967  95.48729 115.48967  95.48729 115.48967  95.48729
##        M7        M7        M7        M7        M7        M7        M7
## 115.48967  95.48729 115.48967  95.48729 115.48967 115.48967  95.48729
##        F2        F2        F2        F2        F2        F2        F2
## 247.15505 267.15743 247.15505 267.15743 247.15505 267.15743 247.15505
##        F2        F2        F2        F2        F2        F2        F2
## 267.15743 247.15505 267.15743 247.15505 267.15743 267.15743 247.15505
##        M3        M3        M3        M3        M3        M3        M3
## 156.14719 176.14957 156.14719 176.14957 156.14719 176.14957 156.14719
##        M3        M3        M3        M3        M3        M3        M3
## 176.14957 156.14719 176.14957 156.14719 176.14957 176.14957 156.14719
## attr(,"label")
## [1] "Fitted values"
```

From the formula we can see that the female in scenario 1 with polite attitude is $\hat{y} = 256.98691 - 20.00238 - \overbrace{13.575831 - 8.418005 \times 10^{-7}}^{\text{random effects}} = 223.40869$