P8131 HW7 yz4184

Yunlin Zhou

```
library(tidyverse)
library(nlme)
library(ggplot2)
library (lattice)
library(patchwork)
library(lme4)
```

```
# import data
pitch_df = read.csv("./HW7-politeness_data.csv")%>%
  drop_na()%>%
  janitor::clean_names()
```

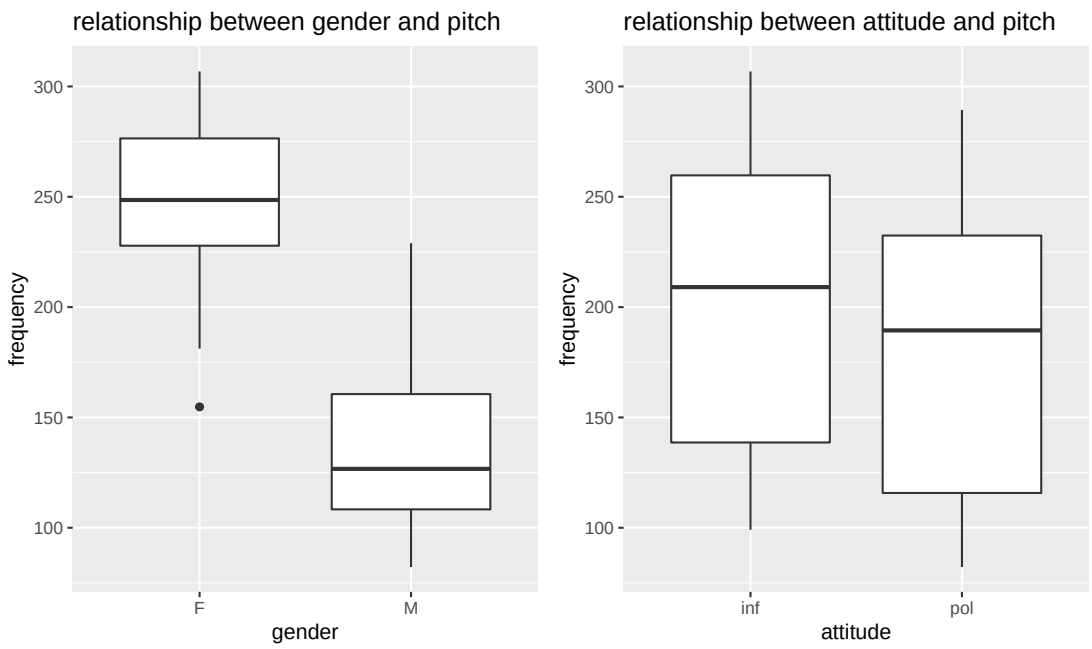
(a) Exploratory analysis

provide boxplots to show the relation between gender/attitude and pitch (ignoring different scenarios).

```
a.p1 = pitch_df %>%
  ggplot(aes(x = gender, y = frequency)) +
  geom_boxplot()+
  labs(title = "relationship between gender and pitch")

a.p2 = pitch_df %>%
  ggplot(aes(x = attitude, y = frequency)) +
  geom_boxplot()+
  labs(title = "relationship between attitude and pitch")

(a.p1 + a.p2)
```



As the plot shows above, we can conclude that female and informal attitude are tending to have higher frequency.

(b)

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

```
LMM1 <- lme (frequency ~ gender + attitude, random = ~1 | subject,  data = pitch_df, method='REML')
summary (LMM1)
```

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

What is the covariance matrix for a subject Yi?

```
VarCorr(LMM1)
```

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

```
VarCorr(LMM1)[1]
```

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

```
VarCorr(LMM1)[2]
```

```
## [1] "847.7049"
```

sigma_b^2 = 598.1953

sigma^2 = 847.7049

sigma_b^2 + sigma^2 = 598.1953 + 847.7049 = 1445.9

For every subject, we have 14 measurements, so the covariance matrix is a 14*14 matrix.

What is the covariance matrix for the estimates of fixed effects?

```
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
```

What are the BLUPs for subject-specific intercepts?

```
random.effects(LMM1)
```

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

What are the residuals?

```
pitch_df$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"
```

(c)

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

```
LMM2 <- lme (frequency ~ gender + attitude + gender * attitude, random = ~1 | subject,  data = pitch_df
summary (LMM2)
```

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

Use likelihood ratio test to compare this model with the model in part (b) to determine whether the interaction term is significantly associated with pitch.

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

LMM.2 <- lme (frequency ~ gender + attitude + gender * attitude, random = ~1 | subject,  data = pitch_d

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
```

Since the p-value for Likelihood ratio test is $0.2392 > 0.05$, we fail to reject the null hypothesis and conclude that the interaction term is not significantly associated with pitch, at the significance level of 0.05.

(d)

Write out the mixed effects model with random intercepts for both subjects and scenarios (gender and attitude being the fixed effects).

Fit the model using lmer in the lme4 package.

```
LMM3 = lmer(frequency ~ gender + attitude + (1 | subject) +  (1 | scenario), data = pitch_df)
summary(LMM3)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: frequency ~ gender + attitude + (1 | subject) + (1 | scenario)
##    Data: pitch_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
```

Write out the covariance matrix for a subject Yi.

```
VarCorr(LMM3)
```

```
##  Groups   Name        Std.Dev.
##  scenario (Intercept) 14.983
##  subject  (Intercept) 24.763
##  Residual             25.254
```

```
VarCorr(LMM3)[1]
```

```
## $scenario
##             (Intercept)
## (Intercept)    224.4994
## attr(,"stddev")
## (Intercept)
##     14.9833
## attr(,"correlation")
##             (Intercept)
## (Intercept)           1
```

```
VarCorr(LMM3)[2]
```

```
## $subject
##             (Intercept)
## (Intercept)    613.1903
## attr(,"stddev")
## (Intercept)
##    24.76268
## attr(,"correlation")
##             (Intercept)
## (Intercept)           1
```

sigma_b^2 = 224.4994 + 613.1903 = 837.6897

sigma^2 = 25.254^2 = 637.7645

sigma_b^2 + sigma^2 = 837.6897 + 637.7645 = 1475.454

What is the interpretation of the coefficient for the fixed effect term attitude?

```
fixed.effects(LMM3)
```

```
## (Intercept)     genderM attitudepol
##   256.98690  -108.79762   -20.00238
```

When the gender is fixed, the mean frequency of polite attitude will be 20.00238 units lower than informal attitude on average.