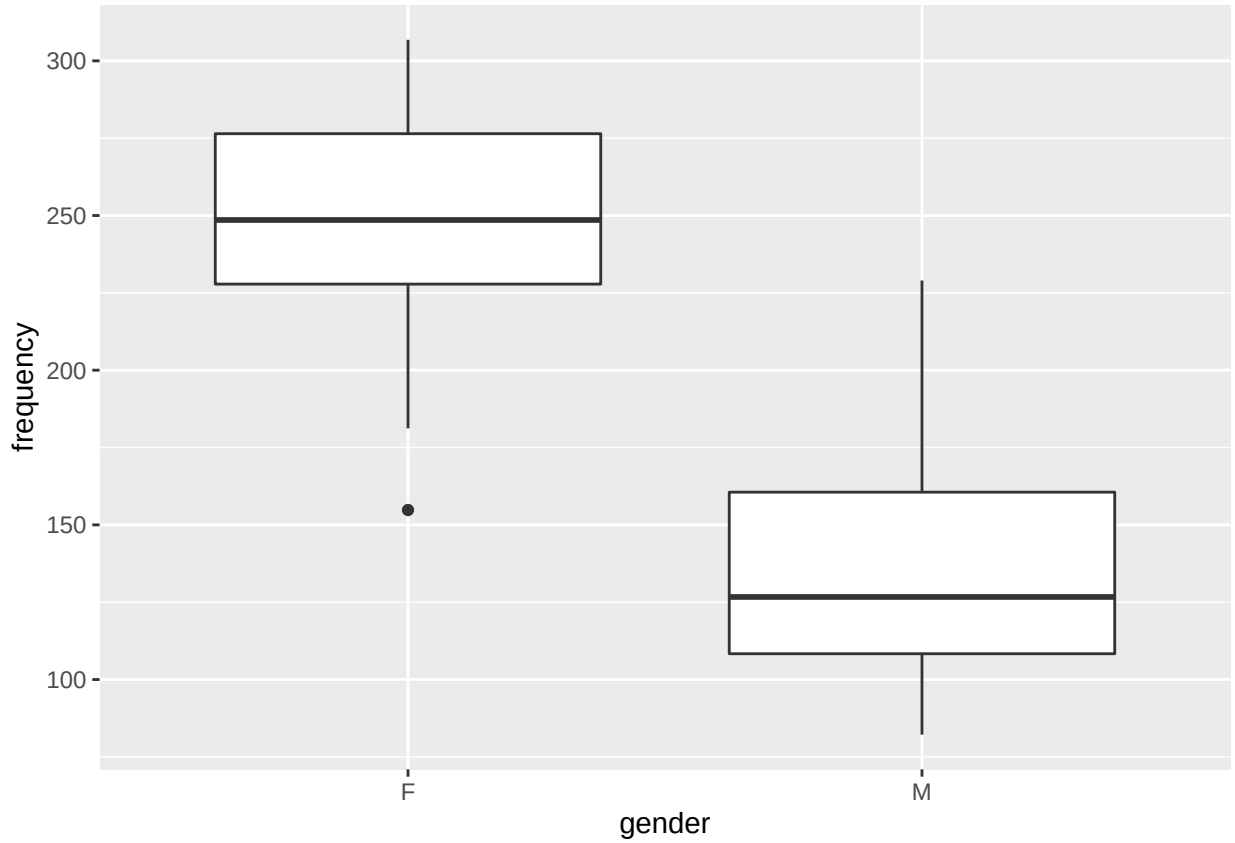
P8131 HW7

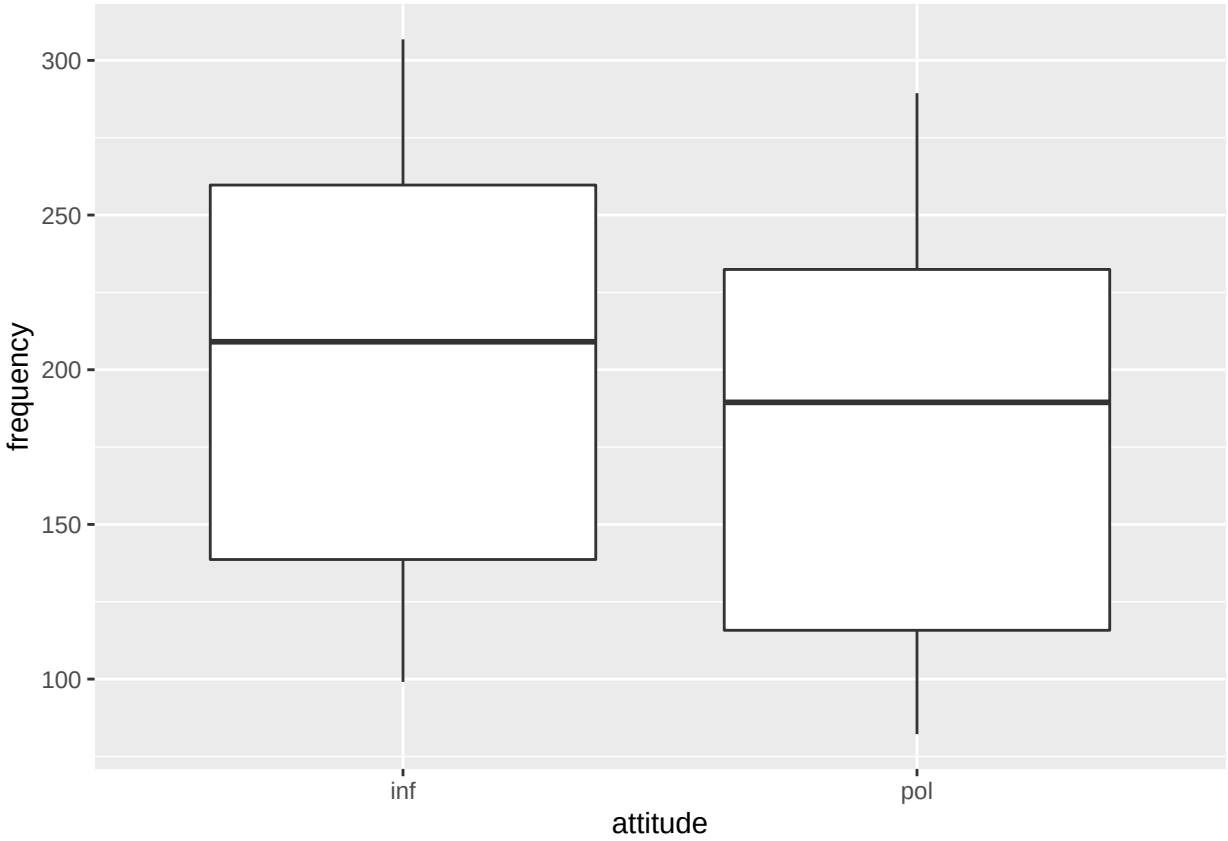
1. The relationship between pitch and politeness

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

```
df <- read_csv("HW7-politeness_data.csv", col_names = TRUE)
# gender v.s. frequency
df %>%
  ggplot(aes(x = gender, y = frequency)) +
  geom_boxplot()
```




```
# attitudes v.s. frequency
df %>%
  ggplot(aes(x = attitude, y = frequency)) +
  geom_boxplot()
```



The distributions of pitch (Hz) by gender are different. With female subjects, the pitch has a average of 250 Hz, and with male subjects the average pitch is only about 125 Hz.The distributions of pitch (Hz) by attitude are also different. Informal registers tend to have a higher average pitch compares with that of formal registers.

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

Fit the model with random intercepts

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

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

$$Pitch_{ij} = 256.987 - 108.798(Gender = M) - 20.002(Attitude = Pol) + b_{i1} + b_{i2} + \epsilon_{ij}$$

Covariance matrix for a subject Y_i

```
VarCorr(LMM.rI) # covariance estimates for random effects and variance for residuals
```

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

Therefore variance of residuals is $\sigma^2 = 847.7049$, and the subject-specific variance for random effects is $\sigma_b^2 = 598.1953$. And there are 14 observations for each subject. Therefore $\sigma^2 + \sigma_b^2 = 1445.9$ So the covariance matrix becomes:

$$\begin{bmatrix} 1445.9 & 598.1953 & ... & 598.1953 \\ 598.1953 & 1445.9 & ... & 598.1953 \\ 598.1953 & 598.1953 & ... & 598.1953 \\ ... & & & \\ 598.1953 & 598.1953 & ... & 598.1953 \end{bmatrix}_{14 \cdot 14}$$

Covariance matrix for the estimates of fixed effects

```
vcov(LMM.rI) # 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
```

```
#
#fixed.effects(LMM.rI) # fixed effects coeff
```

BLUP for Subject i and Residuals

```
# ordered random effects, BLUP (in this case, just b_i)
random.effects(LMM.rI)
```

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

```
# fixed+random residuals
LMM.rI$residuals
```

```
##          fixed     subject
## 1  -23.6845238 -10.1086926
## 2  -52.4869048 -38.9110735
## 3   48.1154762  61.6913074
## 4    2.7130952  16.2889265
## 5  -33.0845238 -19.5086926
## 6   29.9130952  43.4889265
## 7   13.8154762  27.3913074
## 8   19.8130952  33.3889265
## 9   -5.0845238   8.4913074
## 10  -4.5869048   8.9889265
## 11 -55.7845238 -42.2086926
## 12 -26.2869048 -12.7110735
## 13 -40.4869048 -26.9110735
## 14 -82.1845238 -68.6086926
## 15  -7.2845238 -10.6898326
## 16 -19.6869048 -23.0922136
## 17  -0.1845238  -3.5898326
## 18  -5.9869048  -9.3922136
## 19  30.0154762  26.6101674
## 20   9.0130952   5.6077864
## 21  38.4154762  35.0101674
## 22  49.8130952  46.4077864
## 23  -4.3845238  -7.7898326
## 24  -4.4869048  -7.8922136
## 25 -10.4845238 -13.8898326
## 26  21.8130952  18.4077864
## 27   7.4130952   4.0077864
## 28 -51.4845238 -54.8898326
## 29 -17.4869048 -22.2262298
## 30 -24.5892857 -29.3286108
## 31 100.8130952  96.0737702
```

```
## 32 -33.2892857 -38.0286108
## 33 -15.9869048 -20.7262298
## 34  65.4107143  60.6713892
## 35  65.2130952  60.4737702
## 36  14.7107143   9.9713892
## 37 -26.3869048 -31.1262298
## 38 -21.2892857 -26.0286108
## 39 -18.1869048 -22.9262298
## 40 -11.9892857 -16.7286108
## 41  -2.1892857  -6.9286108
## 42  -1.6869048  -6.4262298
## 43 -42.0869048  -9.3872916
## 44 -49.0892857 -16.3896725
## 45 -45.9869048 -13.2872916
## 46 -43.8892857 -11.1896725
## 47 -42.2869048  -9.5872916
## 48 -37.9892857  -5.2896725
## 49 -31.0869048   1.6127084
## 50 -28.1892857   4.5103275
## 51 -34.4869048  -1.7872916
## 52 -45.2892857 -12.5896725
## 53 -19.3869048  13.3127084
## 54 -39.9892857  -7.2896725
## 55 -23.7892857   8.9103275
## 56 -20.5869048  12.1127084
## 57  -4.2845238 -14.4550462
## 58 -25.6869048 -35.8574271
## 59   9.3154762  -0.8550462
## 60   2.7130952  -7.4574271
## 61  52.4154762  42.2449538
## 62  44.8130952  34.6425729
## 63   6.2154762  -3.9550462
## 64  39.2130952  29.0425729
## 65  40.7154762  30.5449538
## 66  37.2130952  27.0425729
## 67 -28.9845238 -39.1550462
## 68 -31.0869048 -41.2574271
## 69  24.0130952  13.8425729
## 70  -9.7845238 -19.9550462
## 71  25.6130952  -2.3471929
## 72  40.6107143  12.6504261
## 73  14.2130952 -13.7471929
## 74  51.5107143  23.5504261
## 75  32.0130952   4.0528071
## 76  37.9107143   9.9504261
## 77  79.3130952  51.3528071
## 78  42.7107143  14.7504261
## 79  32.5130952   4.5528071
## 80   8.3107143 -19.6495739
## 81  18.5130952  -9.4471929
## 82   9.8107143 -18.1495739
## 83  12.9107143 -15.0495739
## 84  25.1130952  -2.8471929
## attr(,"std")
```

```
##  [1] 29.11537 29.11537 29.11537 29.11537 29.11537 29.11537 29.11537 29.11537
##  [9] 29.11537 29.11537 29.11537 29.11537 29.11537 29.11537 29.11537 29.11537
## [17] 29.11537 29.11537 29.11537 29.11537 29.11537 29.11537 29.11537 29.11537
## [25] 29.11537 29.11537 29.11537 29.11537 29.11537 29.11537 29.11537 29.11537
## [33] 29.11537 29.11537 29.11537 29.11537 29.11537 29.11537 29.11537 29.11537
## [41] 29.11537 29.11537 29.11537 29.11537 29.11537 29.11537 29.11537 29.11537
## [49] 29.11537 29.11537 29.11537 29.11537 29.11537 29.11537 29.11537 29.11537
## [57] 29.11537 29.11537 29.11537 29.11537 29.11537 29.11537 29.11537 29.11537
## [65] 29.11537 29.11537 29.11537 29.11537 29.11537 29.11537 29.11537 29.11537
## [73] 29.11537 29.11537 29.11537 29.11537 29.11537 29.11537 29.11537 29.11537
## [81] 29.11537 29.11537 29.11537 29.11537
```

(c) Fit a mixed effects model with intercepts for different subjects

H_0 : Model 2 (larger model with the interaction term) is no better than Model 1 (smaller model)

H_1 : Model 2 (larger model with the interaction term) has better performance than Model 1 (smaller model)

```
# do NOT use REML for likelihood ratio
LMM.2 <- lme(frequency ~ gender + attitude + gender*attitude, random = ~1 | subject,  data = df, method
summary(LMM.2)
```

```
## Linear mixed-effects model fit by maximum likelihood
##   Data: 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 ~ gender + attitude + gender * attitude
##                           Value Std.Error DF   t-value p-value
## (Intercept)          260.68571 13.200754 76 19.747790  0.0000
## genderM             -116.19524 18.668685  4 -6.224072  0.0034
## attitudepol          -27.40000  9.066991 76 -3.021951  0.0034
## genderM:attitudepol   14.79524 12.822662 76  1.153835  0.2522
##  Correlation:
##                      (Intr) gendrM atttdp
## genderM              -0.707
## attitudepol          -0.343  0.243
## genderM:attitudepol  0.243 -0.343 -0.707
##
## 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
```

```r
LMM.1 <- lme(frequency ~ gender + attitude, random = ~1 | subject,  data = df, method = 'ML')
# Compare
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 p-value of the test is 0.2392, so we fail to reject the null hypothesis. We conclude that the interaction term is not significantly associated with pitch.

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

Fit the model with random intercepts

```r
# grouped data
LMM.3 <- lme4::lmer(frequency ~ gender + attitude + (1|subject) + (1|scenario),
                    data = df, REML = TRUE)
summary(LMM.3)
```

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

The mixed effect model is

$$Pitch_{ijk} = 256.987 - 108.798(Gender = M)_i - 20.002(Attitude = Pol)_{ij} + b_{ji} + b_{jk} + \epsilon_{ijk}$$

where b_{0i} and b_{0k} are random effect terms for `subject` and `scenario`, and ϵ_{ijk} is the residual. Therefore $i \in F1, F2, F3, M3, M4, M7$, $j \in 0, 1$, and $k \in 1, 2, .., 7$. Note that b_{ji} is the same for all j's, and b_{jk} is the same for all j's.

Covariance matrix for a subject Y_i

The variance of residuals is $\sigma^2 = 637.8$, and the subject-specific variance for random effects of group `subject` is $\sigma^2_{b_{ji}} = 613.2$, and that of group `scenario` $\sigma^2_{b_{jk}} = 224.5$.

The variance of Y_i is therefore:

$$\begin{bmatrix} A_1 & A_2 \\ A_2 & A_1 \end{bmatrix}_{14 \cdot 14}$$

where A_1 and A_2 are symmetric $7 \cdot 7$ matrices:

$$A_1 = \begin{bmatrix} Var[Y_{ijk}] & Cov(Y_{ijk}, Y_{ijm}) & Cov(Y_{ijk}, Y_{ijm}) & ... & Cov(Y_{ijk}, Y_{ijm}) \\ Cov(Y_{ijk}, Y_{ijm}) & Var[Y_{ijk}] & Cov(Y_{ijk}, Y_{ijm}) & ... & Cov(Y_{ijk}, Y_{ijm}) \\ ... & & & & \\ ... & & & Cov(Y_{ijk}, Y_{ijm}) & Var[Y_{ijk}] \end{bmatrix}_{7 \cdot 7}$$

$$A_2 = \begin{bmatrix} Cov(Y_{ijk}, Y_{ink}) & Cov(Y_{ijk}, Y_{inm}) & Cov(Y_{ijk}, Y_{inm}) & ... & Cov(Y_{ijk}, Y_{inm}) \\ Cov(Y_{ijk}, Y_{inm}) & Cov(Y_{ijk}, Y_{ink}) & Cov(Y_{ijk}, Y_{inm}) & ... & Cov(Y_{ijk}, Y_{inm}) \\ ... & & & & \\ ... & & & Cov(Y_{ijk}, Y_{inm}) & Cov(Y_{ijk}, Y_{ink}) \end{bmatrix}_{7 \cdot 7}$$

The values are calculated as:

$$Var[Y_{ijk}] = E[(Y_{ijk} - \mu)^2] = E[(b_{ji} + b_{jk} + \epsilon_{ijk})^2] = \sigma^2 + \sigma^2_{b_{ji}} + \sigma^2_{b_{jk}} = 1475.5$$

$$Cov(Y_{ijk}, Y_{ijm}) = E[(b_{ji} + b_{jk} + \epsilon_{ijk})(b_{ji} + b_{jm} + \epsilon_{ijm})] = \sigma^2_{b_{ji}} = 613.2$$

$$Cov(Y_{ijk}, Y_{ink}) = E[(b_{ji} + b_{jk} + \epsilon_{ijk})(b_{ni} + b_{nk} + \epsilon_{ink})] = \sigma^2_{b_{ji}} + \sigma^2_{b_{jk}} = 837.7$$

$$Cov(Y_{ijk}, Y_{inm}) = E[(b_{ji} + b_{jk} + \epsilon_{ijk})(b_{ni} + b_{nm} + \epsilon_{inm})] = \sigma^2_{b_{ji}} = 613.2$$

Interpretation of attitude coefficient

With the gender of the subject known, formal registers have a lower average pitch of 20.002 Hz comparing with the average pitch of informal register.