P8131 HW7

Brian Jo Hsuan Lee

4/4/2022

Load packages

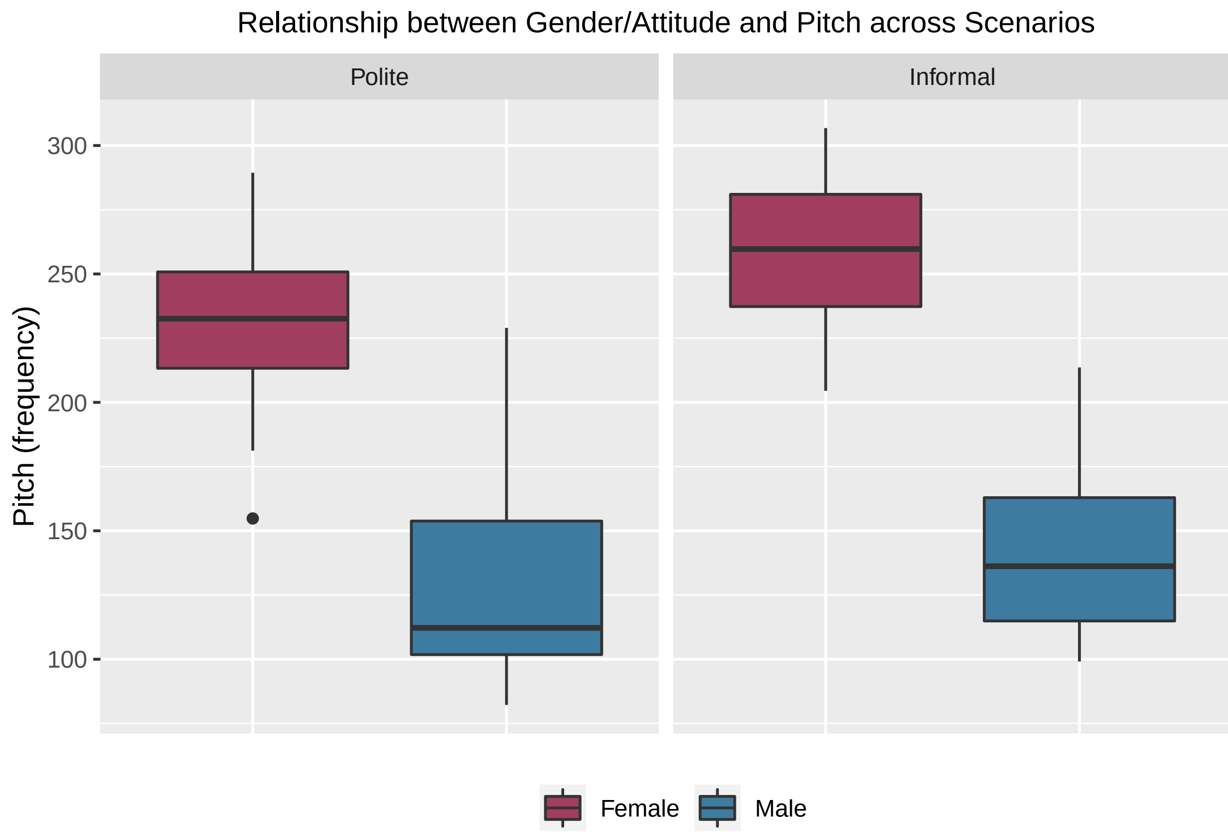
```
library(tidyverse)
library(knitr)
library(nlme)
library(lme4)
```

Import data

```
data = read_csv("HW7-politeness_data.csv", col_types = "ffffd")
```

a) **EDA**

```
data %>%
  mutate(
    attitude = factor(attitude, labels = c("Polite", "Informal"))
  ) %>%
  ggplot(aes(x = gender, y = frequency, fill = gender)) +
  geom_boxplot() +
  facet_grid(cols = vars(attitude)) +
  scale_fill_manual(labels = c("Female", "Male"), values = c("#A13E60", "#3E7BA1")) +
  labs(
    title = "Relationship between Gender/Attitude and Pitch across Scenarios",
    y = "Pitch (frequency)"
  ) +
  theme(
    plot.title = element_text(size = 11, hjust = 0.5),
    axis.title.x = element_blank(),
    axis.text.x = element_blank(),
    axis.ticks.x = element_blank(),
    legend.position = "bottom",
    legend.title = element_blank()
  )
```


Relationship between Gender/Attitude and Pitch across Scenarios
Polite
Informal
Pitch (frequency)
300
250
200
150
100
Female
Male

b) **Fit and interpret a random intercept model for different subjects**

```
# fit a mixed effect model with estimates chosen to optimize the maximum log-likelihood criterion
lmm1 = lme (frequency ~ gender + attitude, random = ~1 | subject, data = data, method ='ML')
```

The covariance matrix for the pitch `frequency` i of a particular subject is composed of the marginal variances of each population-shared predictor as its diagonal, and the marginal covariances of any two of those predictors in their corresponding entries. The diagonals are all equal, and the non-diagonal entries are all equal for a linear mixed effect model.

In this random intercept model $Y_{ij} = (\beta_o + b_i) + X_{ij}^T \beta + \epsilon_{ij}$, Y_{ij} and X_{ij} are the estimated i^{th} `frequency` and its vector of predictors, `genderM` and `attitudeM`, in condition j of a particular subject.
$b_i \sim N(0, \sigma_b^2)$ is the random `subject`-specific intercept effect for the i^{th} `frequency`, and
$\epsilon_{ij} \sim N(0, \sigma_b)$ is the within-`subject` error at condition j for the i^{th} `frequency`.
Note b_i and ϵ_{ij} are independent, i.e. $\text{cov}(b_i, \epsilon_{ij}) = 0$, $\text{cov}(\epsilon_{im}, \epsilon_{in}) = 0$.

The covariance matrix for `frequency` i of a subject is derived with equations
$cov(Y_{im}, Y_{in}) = cov(b_i + \epsilon_{im}, b_i + \epsilon_{in}) = cov(b_i, b_i) + cov(b_i, \epsilon_{in}) + cov(\epsilon_{im}, b_i) + cov(\epsilon_{im}, \epsilon_{in})$
$= Var[b_i] + 0 + 0 + 0 = \sigma_b^2$
for the marginal covariance between `frequency` i pairs under conditions m and n , and
$Var[Y_{ij}] = Var[b_i + \epsilon_{ij}] = Var[b_i] + Var[\epsilon_{ij}] = \sigma_b^2 + \sigma^2$

```
# obtain the random subject-specific covariance estimate (sigma^2_b)
randeff_cov = as.double(VarCorr(lmm1)[1,1])

# obtain the random population-shared residual variance estimate (sigma^2)
res_var = as.double(VarCorr(lmm1)[2,1])

# build the covariance matrix for a particular subject with the estimates
# where the marginal variance for the subject is the sum of the two values
pop_pred = c("genderM", "attitudeinf")
cov_y =
  matrix(
    rep(randeff_cov, length(pop_pred)^2),
    nrow = length(pop_pred),
    dimnames = list(pop_pred, pop_pred)
  )
diag(cov_y) = randeff_cov + res_var

kable(cov_y, "simple")
```

	genderM	attitudeinf
genderM	1216.2266	379.3897
attitudeinf	379.3897	1216.2266

The covariance matrix for the fixed effect estimates

```
kable(vcov(lmm1), "simple")
```

	(Intercept)	genderM	attitudeinf
(Intercept)	156.35027	-146.3879	-19.92469
genderM	-146.38793	292.7759	0.00000
attitudeinf	-19.92469	0.0000	39.84938

```r
# # or alternatively ...
# lmm1$varFix
```

BLUPs for subject-specific intercepts, which are the random effect coefficients

```r
kable(random.effects(lmm1), "simple")
```

	(Intercept)
F1	-12.915173
F3	3.239592
M4	4.508689
M7	-31.108310
F2	9.675581
M3	26.599621

Residuals (is there a better way to show the residuals?)

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

```
##           F1          F1          F1          F1          F1          F1
## -10.76935066 -39.57173161  61.03064934  15.62826839 -20.16935066  42.82826839
##           F1          F1          F1          F1          F1          F1
##  26.73064934  32.72826839   7.83064934   8.32826839 -42.86935066 -13.37173161
##           F1          F1          F3          F3          F3          F3
## -27.57173161 -69.26935066 -10.52411574 -22.92649669  -3.42411574  -9.22649669
##           F3          F3          F3          F3          F3          F3
##  26.77588426   5.77350331  35.17588426  46.57350331  -7.62411574  -7.72649669
##           F3          F3          F3          F3          M4          M4
## -13.72411574  18.57350331   4.17350331 -54.72411574 -21.99559397 -29.09797492
##           M4          M4          M4          M4          M4          M4
##  96.30440603 -37.79797492 -20.49559397  60.90202508  60.70440603  10.20202508
##           M4          M4          M4          M4          M4          M4
## -30.89559397 -25.79797492 -22.69559397 -16.49797492  -6.69797492  -6.19559397
##           M7          M7          M7          M7          M7          M7
## -10.97859473 -17.98097568 -14.87859473 -12.78097568 -11.17859473  -6.88097568
##           M7          M7          M7          M7          M7          M7
##   0.02140527   2.91902432  -3.37859473 -14.18097568  11.72140527  -8.88097568
##           M7          M7          F2          F2          F2          F2
##   7.31902432  10.52140527 -13.96010503 -35.36248598  -0.36010503  -6.96248598
##           F2          F2          F2          F2          F2          F2
##  42.73989497  35.13751402  -3.46010503  29.53751402  31.03989497  27.53751402
##           F2          F2          F2          F2          M3          M3
## -38.66010503 -40.76248598  14.33751402 -19.46010503  -0.98652558  14.01109346
##           M3          M3          M3          M3          M3          M3
## -12.38652558  24.91109346   5.41347442  11.31109346  52.71347442  16.11109346
##           M3          M3          M3          M3          M3          M3
##   5.91347442 -18.28890654  -8.08652558 -16.78890654 -13.68890654  -1.48652558
## attr(,"label")
## [1] "Fitted values"
```

c) **Fit a similar random intercept model - but with an interaction term - and compare it with the first model**

```
# fit a mixed effect model, also with estimates chosen to optimize the maximum log-likelihood criterion
lmm2 = lme (frequency ~ gender * attitude, random = ~1 | subject, data = data, method ='ML')

# compare it with the first model
lmm1_lmm2_pval = anova(lmm2, lmm1)[2, 9]
ifelse(lmm1_lmm2_pval < 0.05,
       "Reject the null hypothesis and suggest the new model with the interaction term has a better fit
       "Fail to reject the null hypothesis and suggest the inclusion of the interaction term does not i
```

```
## [1] "Fail to reject the null hypothesis and suggest the inclusion of the interaction term does not i
```

After comparing the 2 models using the likelihood ratio test, it is concluded that the interaction term for gender and attitude does not create a better fit for modeling pitch, and therefore it is not significantly associated with pitch.

d) **Fit and interpret a random intercept model for different subjects and scenarios**

```
# fit a mixed effect model, again, with estimates chosen to optimize the maximum log-likelihood criteri
lmm3 = lmer(frequency ~ gender + attitude + (1 | subject) + (1 | scenario), data = data, REML = F)
```

As before, the covariance matrix for `frequency` i of a particular subject in a scenario is composed of the marginal variances of each population-shared predictor and the marginal covariances of any two of those predictors.

In this random intercept model $Y_{ij} = (\beta_o + b_{sub,i} + b_{sce,i}) + X_{ij}^T \beta + \epsilon_{ij}$, Y_{ij} and X_{ij} are the estimated i^{th} `frequency`
and its vector of predictors, `genderM` and `attitudeM`, in condition j of a particular subject and scenario.
$b_{sub,i} \sim N(0, g_{sub})$ is the random `subject`-specific intercept effect for the i^{th} `frequency`,
$b_{sce,i} \sim N(0, g_{sce})$ is the random `scenario`-specific intercept effect for the i^{th} `frequency`,
$\epsilon_{ij} \sim N(0, \sigma_b)$ is the within-`subject-scenario` error at condition j for the i^{th} `frequency`.
Note $b_{sub,i}$, $b_{sce,i}$ and ϵ_{ij} are independent, i.e. $cov(b_{sub,i}, b_{sce,i}) = 0$.

The covariance matrix for `frequency` i of a particular subject and scenario is derived with equations
$cov(Y_{im}, Y_{in}) = cov(b_{sub,i} + b_{sce,i} + \epsilon_{im}, b_{sub,i} + b_{sce,i} + \epsilon_{in}) = cov(b_{sub,i} + b_{sce,i}, b_{sub,i} + b_{sce,i})$
$= cov(b_{sub,i}, b_{sub,i}) + cov(b_{sce,i}, b_{sub,i}) + cov(b_{sub,i}, b_{sce,i}) + cov(b_{sce,i}, b_{sce,i})$
$= Var[b_{sub,i}] + 0 + 0 + Var[b_{sce,i}] = g_{sub} + g_{sce}$
for the marginal covariance between `frequency` i pairs under conditions m and n, and
$Var[Y_{ij}] = Var[b_{sub,i} + b_{sce,i} + \epsilon_{ij}] = Var[b_{sub,i}] + Var[b_{sce,i}] + Var[\epsilon_{ij}] = g_{sub} + g_{sce} + \sigma^2$

```
cov_obj = as.data.frame(VarCorr(lmm3))

# obtain the residual variance estimate (sigma^2)
res_var2 = cov_obj[3,4]

# obtain the subject covariance estimate (sigma^2_bsub)
sub_cov = cov_obj[2,4]

# obtain the scenario covariance estimate (sigma^2_bsce)
sce_cov = cov_obj[1,4]

# build a covariance matrix with the covariance and variance estimates
# where the variance for Y is the sum of the two values
cov_y2 =
  matrix(
    rep(sub_cov + sce_cov, length(pop_pred)^2),
```

```
    nrow = length(pop_pred),
    dimnames = list(pop_pred, pop_pred)
  )
diag(cov_y2) = sub_cov + sce_cov + res_var2

kable(cov_y2, "simple")
```

	genderM	attitudeinf
genderM	1255.0578	625.5026
attitudeinf	625.5026	1255.0578

Acquire the fixed effect coefficients

```
kable(fixed.effects(lmm3), "simple")
```

	x
(Intercept)	236.98452
genderM	-108.79762
attitudeinf	20.00238

The fixed effect `attitude` is a categorical variable, so the coefficient for `attitudeinf` is the relative change in pitch when the attitude switches from polite to informal while adjusting for gender. That is to say, when the attitude is informal, the pitch frequency increases by 20 units for any subject in any scenario.