Assignment 10

Sachin Shubham

4/3/2021

1 Task

You got a new assignment in the RD department. You need to build a model to predict the voice pitch based on the collected data. Upload the data file politeness data.csv.The difference in politeness level is represented in the column called “attitude”. In that column, “pol” stands for polite and “inf” for informal. Sex is represented as “F” and “M” in the column “gender”. The dependent measure is “frequency”, which is the voice pitch measured in Hertz (Hz). To remind you,higher values mean higher pitch.

0. Explore the data set. Produce a box plot representing the relationship between politeness and pitch (frequency attitude*gender). Comment about observed picture.

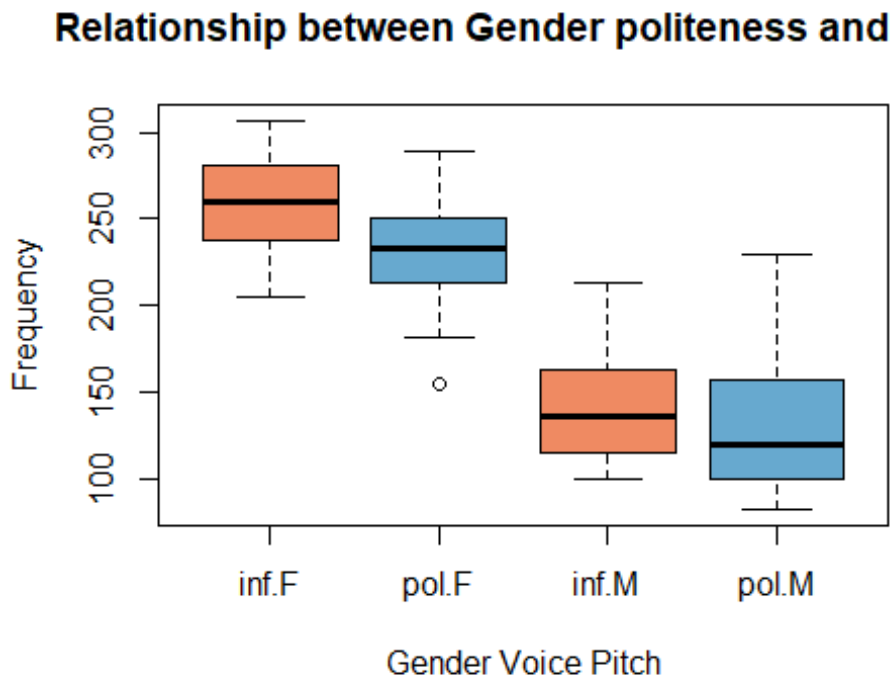
Build following models to predict the frequency:

1. Model that used the fixed effect “attitude” (polite vs. informal) to predict voice pitch, controlling for by-subject and by-item variability.
2. Model that used the fixed effect “gender” (male vs. female) to predict voice pitch, controlling for by-subject and by-item variability.
3. Model that used the both “gender” and ”attitude” as fixed effect to predict voice pitch, controlling for by-subject and by-item variability.
4. Comment on the summary of the models. Which model should be used for the most accurate prediction?

Solution:

```
library(lme4)

## Warning: package 'lme4' was built under R version 4.0.4

## Loading required package: Matrix

library(arm)

## Warning: package 'arm' was built under R version 4.0.4

## Loading required package: MASS

##
## arm (Version 1.11-2, built: 2020-7-27)

## Working directory is D:/Statistical Methods/Assignments/Assignment 8
```

```
library(Matrix)

politeness_data<-read.csv(file.choose())


#0

boxplot(frequency ~ attitude*gender, data=politeness_data,
        col=c("#ef8a62","#67a9cf"),
        main="Relationship between Gender politeness and pitch",
        xlab="Gender Voice Pitch", ylab="Frequency")
```



As per result from Boxplot

(1) The median is lower for the polite than for the informal attitude

(2) There is more overlap between the informal and polite attitude for the males than for the females.

```
#1

lmer_attitude = lmer(frequency ~ attitude  + (1|subject) + (1|scenario),
data=politeness_data)
display(lmer_attitude)
```

```
## lmer(formula = frequency ~ attitude + (1 | subject) + (1 | scenario),
##     data = politeness_data)
##             coef.est coef.se
## (Intercept) 202.59    26.75
## attitudepol -19.69     5.58
##
## Error terms:
##  Groups   Name        Std.Dev.
##  scenario (Intercept) 14.80
##  subject  (Intercept) 63.36
##  Residual             25.42
## ---
## number of obs: 83, groups: scenario, 7; subject, 6
## AIC = 803.5, DIC = 820.6
## deviance = 807.0

#2

lmer_gender <- lmer(frequency ~ gender + (1|subject) + (1|scenario),
data=politeness_data)
display(lmer_gender)

## lmer(formula = frequency ~ gender + (1 | subject) + (1 | scenario),
##     data = politeness_data)
##             coef.est coef.se
## (Intercept)  246.99    15.84
## genderM     -108.24    21.06
##
## Error terms:
##  Groups   Name        Std.Dev.
##  scenario (Intercept) 14.35
##  subject  (Intercept) 24.71
##  Residual             27.40
## ---
## number of obs: 83, groups: scenario, 7; subject, 6
## AIC = 802.4, DIC = 821.1
## deviance = 806.7

#3

lmer_attitude_gender <- lmer(frequency ~ gender + attitude + (1|subject) +
(1|scenario), data=politeness_data)
display(lmer_attitude_gender)

## lmer(formula = frequency ~ gender + attitude + (1 | subject) +
##     (1 | scenario), data = politeness_data)
##             coef.est coef.se
## (Intercept)  256.85    16.12
## genderM     -108.52    21.01
## attitudepol  -19.72     5.58
##
```

```
## Error terms:
##  Groups   Name        Std.Dev.
##  scenario (Intercept) 14.81
##  subject  (Intercept) 24.81
##  Residual             25.41
## ---
## number of obs: 83, groups: scenario, 7; subject, 6
## AIC = 787.5, DIC = 814.7
## deviance = 795.1
```

```
#4

AIC(lmer_attitude, lmer_gender, lmer_attitude_gender)
```

```
##                       df      AIC
## lmer_attitude          5 803.4536
## lmer_gender            5 802.3546
## lmer_attitude_gender   6 787.4547
```

```
BIC(lmer_attitude, lmer_gender, lmer_attitude_gender)
```

```
##                       df      BIC
## lmer_attitude          5 815.5479
## lmer_gender            5 814.4488
## lmer_attitude_gender   6 801.9678
```

As we know that the smaller the AIC or BIC, the better the fit.

comparing the AIC of the model:

lmer_attitude: 803.4536

lmer_gender 802.3546

lmer_attitude_gender 787.4547

comparing the BIC of the model:

lmer_attitude 815.5479

lmer_gender 814.4488

lmer_attitude_gender 801.9678

We can conclude that model lmer_attitude_gender is more significant than model lmer_attitude and lmer_gender. Therefore the model that used the both “gender” and ”attitude” as fixed effect to predict voice pitch is the the most accurate prediction.