Group-Module-3.R

Duker

2020-12-09

setwd("C:/Users/Duker/Desktop/Fall 2020/CS 614/Class Work/Group Module 3")  
library(lme4)

## Loading required package: Matrix

library(lmerTest)

##   
## Attaching package: 'lmerTest'

## The following object is masked from 'package:lme4':  
##   
## lmer

## The following object is masked from 'package:stats':  
##   
## step

library(MASS)  
  
data = read.csv("politeness\_data.csv")  
  
# Random intercept model with subject as the random effect and gender + attitude as predictor variables.   
model1 = lmer(frequency~1 + (1|gender) + (1|attitude), data = data)  
summary(model1)

## Linear mixed model fit by REML. t-tests use Satterthwaite's method [  
## lmerModLmerTest]  
## Formula: frequency ~ 1 + (1 | gender) + (1 | attitude)  
## Data: data  
##   
## REML criterion at convergence: 830.4  
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -2.34619 -0.73394 -0.07035 0.70048 2.75373   
##   
## Random effects:  
## Groups Name Variance Std.Dev.  
## gender (Intercept) 5834.9 76.39   
## attitude (Intercept) 160.4 12.66   
## Residual 1272.8 35.68   
## Number of obs: 83, groups: gender, 2; attitude, 2  
##   
## Fixed effects:  
## Estimate Std. Error df t value Pr(>|t|)  
## (Intercept) 192.834 54.890 1.054 3.513 0.166

# Random slope model with subject as the random effect, but with attitude as the slope effect. Still  
# have gender + attitude as predictor variables.   
  
model2 = lmerTest::lmer(frequency~ subject + (subject|attitude), data = data,  
 na.action = na.exclude)

## boundary (singular) fit: see ?isSingular

summary(model2)

## Linear mixed model fit by REML. t-tests use Satterthwaite's method [  
## lmerModLmerTest]  
## Formula: frequency ~ subject + (subject | attitude)  
## Data: data  
##   
## REML criterion at convergence: 753.9  
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -2.2181 -0.5282 -0.1447 0.4952 2.9019   
##   
## Random effects:  
## Groups Name Variance Std.Dev. Corr   
## attitude (Intercept) 400.1369 20.0034   
## subjectF2 15.8570 3.9821 -1.00   
## subjectF3 0.1145 0.3383 -1.00 1.00   
## subjectM3 76.3123 8.7357 -1.00 1.00 1.00   
## subjectM4 381.7135 19.5375 -1.00 1.00 1.00 1.00   
## subjectM7 91.9164 9.5873 -1.00 1.00 1.00 1.00 1.00  
## Residual 825.1974 28.7262   
## Number of obs: 83, groups: attitude, 2  
##   
## Fixed effects:  
## Estimate Std. Error df t value Pr(>|t|)   
## (Intercept) 232.036 16.094 1.044 14.418 0.0395 \*   
## subjectF2 26.150 11.217 7.472 2.331 0.0503 .   
## subjectF3 18.700 10.860 70.835 1.722 0.0895 .   
## subjectM3 -63.057 12.492 2.282 -5.048 0.0281 \*   
## subjectM4 -86.083 17.700 1.167 -4.864 0.1027   
## subjectM7 -129.857 12.800 2.037 -10.145 0.0090 \*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Correlation of Fixed Effects:  
## (Intr) sbjcF2 sbjcF3 sbjcM3 sbjcM4  
## subjectF2 -0.547   
## subjectF3 -0.357 0.489   
## subjectM3 -0.728 0.545 0.445   
## subjectM4 -0.893 0.493 0.324 0.653   
## subjectM7 -0.752 0.543 0.436 0.631 0.674  
## optimizer (nloptwrap) convergence code: 0 (OK)  
## boundary (singular) fit: see ?isSingular

# Random intercept model with subject and scenario as the random effects and gender + attitude as  
# predictor variables. Still have gender + attitude as predictor variables.  
model3 = lmer(frequency~gender+attitude + (1|scenario) + (1|subject), data = data)  
summary(model3)

## Linear mixed model fit by REML. t-tests use Satterthwaite's method [  
## lmerModLmerTest]  
## Formula: frequency ~ gender + attitude + (1 | scenario) + (1 | subject)  
## Data: data  
##   
## REML criterion at convergence: 775.5  
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -2.2591 -0.6236 -0.0772 0.5388 3.4795   
##   
## Random effects:  
## Groups Name Variance Std.Dev.  
## scenario (Intercept) 219.5 14.81   
## subject (Intercept) 615.6 24.81   
## Residual 645.9 25.41   
## Number of obs: 83, groups: scenario, 7; subject, 6  
##   
## Fixed effects:  
## Estimate Std. Error df t value Pr(>|t|)   
## (Intercept) 256.846 16.116 5.432 15.938 9.06e-06 \*\*\*  
## genderM -108.516 21.013 4.007 -5.164 0.006647 \*\*   
## attitudepol -19.721 5.584 70.054 -3.532 0.000735 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Correlation of Fixed Effects:  
## (Intr) gendrM  
## genderM -0.652   
## attitudepol -0.173 0.004

# Problem 1 ####  
var.ICC = function(sigmaB, sigmaW){  
 return(round(sigmaB/(sigmaB + sigmaW),3))  
}  
  
gender.sigmaB = 5834.9  
sigmaW = 1272.8  
model1.gender.ICC = var.ICC(gender.sigmaB, sigmaW)  
model1.gender.ICC

## [1] 0.821

# Looking at the ICC for gender, we can see that ~82% of the variation is associated within the   
# gender class.  
  
attitude.sigmaB = 160.4  
model1.attitude.ICC = var.ICC(attitude.sigmaB, sigmaW)  
model1.attitude.ICC

## [1] 0.112

# Similarly, the ICC for attitude is only 11%, which implies low variation within that class.  
  
summary(model2)

## Linear mixed model fit by REML. t-tests use Satterthwaite's method [  
## lmerModLmerTest]  
## Formula: frequency ~ subject + (subject | attitude)  
## Data: data  
##   
## REML criterion at convergence: 753.9  
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -2.2181 -0.5282 -0.1447 0.4952 2.9019   
##   
## Random effects:  
## Groups Name Variance Std.Dev. Corr   
## attitude (Intercept) 400.1369 20.0034   
## subjectF2 15.8570 3.9821 -1.00   
## subjectF3 0.1145 0.3383 -1.00 1.00   
## subjectM3 76.3123 8.7357 -1.00 1.00 1.00   
## subjectM4 381.7135 19.5375 -1.00 1.00 1.00 1.00   
## subjectM7 91.9164 9.5873 -1.00 1.00 1.00 1.00 1.00  
## Residual 825.1974 28.7262   
## Number of obs: 83, groups: attitude, 2  
##   
## Fixed effects:  
## Estimate Std. Error df t value Pr(>|t|)   
## (Intercept) 232.036 16.094 1.044 14.418 0.0395 \*   
## subjectF2 26.150 11.217 7.472 2.331 0.0503 .   
## subjectF3 18.700 10.860 70.835 1.722 0.0895 .   
## subjectM3 -63.057 12.492 2.282 -5.048 0.0281 \*   
## subjectM4 -86.083 17.700 1.167 -4.864 0.1027   
## subjectM7 -129.857 12.800 2.037 -10.145 0.0090 \*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Correlation of Fixed Effects:  
## (Intr) sbjcF2 sbjcF3 sbjcM3 sbjcM4  
## subjectF2 -0.547   
## subjectF3 -0.357 0.489   
## subjectM3 -0.728 0.545 0.445   
## subjectM4 -0.893 0.493 0.324 0.653   
## subjectM7 -0.752 0.543 0.436 0.631 0.674  
## optimizer (nloptwrap) convergence code: 0 (OK)  
## boundary (singular) fit: see ?isSingular

# There is statistical significance when looking at the fixed effects of subjectM3, M7 & the Intercept.  
  
summary(model3)

## Linear mixed model fit by REML. t-tests use Satterthwaite's method [  
## lmerModLmerTest]  
## Formula: frequency ~ gender + attitude + (1 | scenario) + (1 | subject)  
## Data: data  
##   
## REML criterion at convergence: 775.5  
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -2.2591 -0.6236 -0.0772 0.5388 3.4795   
##   
## Random effects:  
## Groups Name Variance Std.Dev.  
## scenario (Intercept) 219.5 14.81   
## subject (Intercept) 615.6 24.81   
## Residual 645.9 25.41   
## Number of obs: 83, groups: scenario, 7; subject, 6  
##   
## Fixed effects:  
## Estimate Std. Error df t value Pr(>|t|)   
## (Intercept) 256.846 16.116 5.432 15.938 9.06e-06 \*\*\*  
## genderM -108.516 21.013 4.007 -5.164 0.006647 \*\*   
## attitudepol -19.721 5.584 70.054 -3.532 0.000735 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Correlation of Fixed Effects:  
## (Intr) gendrM  
## genderM -0.652   
## attitudepol -0.173 0.004

# Problem 3 ####  
m1 = lmer(frequency~1 + (1|gender) + (1|attitude), data = data, REML = F)  
m2 = lmerTest::lmer(frequency~ subject + (subject|attitude), data = data,  
 na.action = na.exclude, REML = F)

## boundary (singular) fit: see ?isSingular

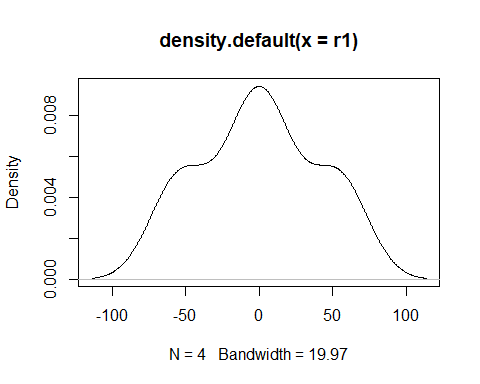
m3 = lmer(frequency~gender+attitude + (1|scenario) + (1|subject), data = data, REML = F)  
  
anova(m1, m2, m3)

## Data: data  
## Models:  
## m1: frequency ~ 1 + (1 | gender) + (1 | attitude)  
## m3: frequency ~ gender + attitude + (1 | scenario) + (1 | subject)  
## m2: frequency ~ subject + (subject | attitude)  
## npar AIC BIC logLik deviance Chisq Df Pr(>Chisq)   
## m1 4 847.91 857.58 -419.95 839.91   
## m3 6 807.10 821.61 -397.55 795.10 44.8051 2 1.865e-10 \*\*\*  
## m2 28 847.34 915.06 -395.67 791.34 3.7657 22 1   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

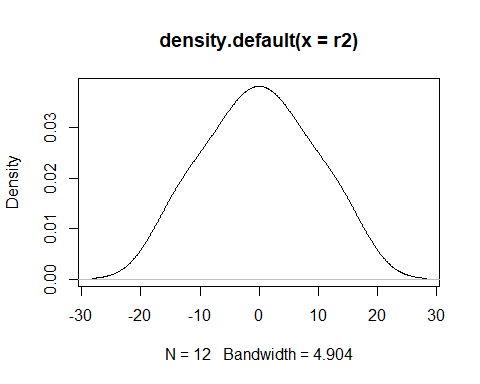
AIC(m1, m2, m3)

## df AIC  
## m1 4 847.9067  
## m2 28 847.3358  
## m3 6 807.1015

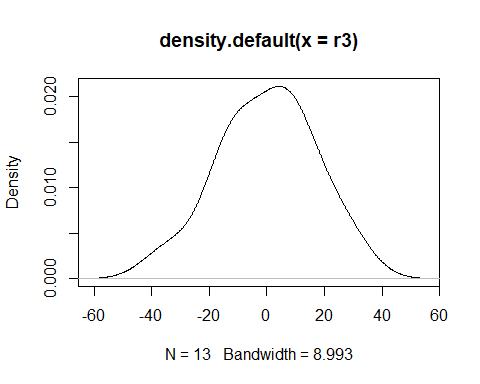
# Model 3 has the AIC & BIC thus it is the simplest model with the best performance.  
  
# Problem 4 ####  
r1 = unlist(ranef(model1))  
plot(density(r1))



r2 = unlist(ranef(model2))  
plot(density(r2))



r3 = unlist(ranef(model3))  
plot(density(r3))



# The plots of the random effects all resemble a bell-shaped curve thus confirming these effects  
# are normally distributed.