# Module 3

Ryan Richardson

11/12/2020

## R Markdown

```
data = read.csv("politeness_data.csv")
data$gender = as.factor(data$gender)
data$attitude = as.factor(data$attitude)
data$scenario = as.factor(data$scenario)
```

## Including Plots

You can also embed plots, for example:

```
randomIntercept = lmer(frequency ~ (1|subject) + gender + attitude, data = data)

randomSlope =  lmer(frequency ~ (attitude|subject) + gender + attitude, data = data)
```

```
## boundary (singular) fit: see ?isSingular
```

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

### Random Intercept Model

```
summary(randomIntercept)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: frequency ~ (1 | subject) + gender + attitude
##    Data: data
##
## REML criterion at convergence: 786.7
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.3619 -0.5305 -0.1724  0.4647  3.2260
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  subject  (Intercept) 603.9    24.57
##  Residual             850.9    29.17
## Number of obs: 83, groups:  subject, 6
##
## Fixed effects:
##             Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)  256.691     15.226   4.378  16.859 3.78e-05 ***
```

```
## genderM     -108.205     21.063    4.009  -5.137  0.00677 **
## attitudepol  -19.410      6.407   76.018  -3.030  0.00334 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) gendrM
## genderM     -0.692
## attitudepol -0.210  0.004
```

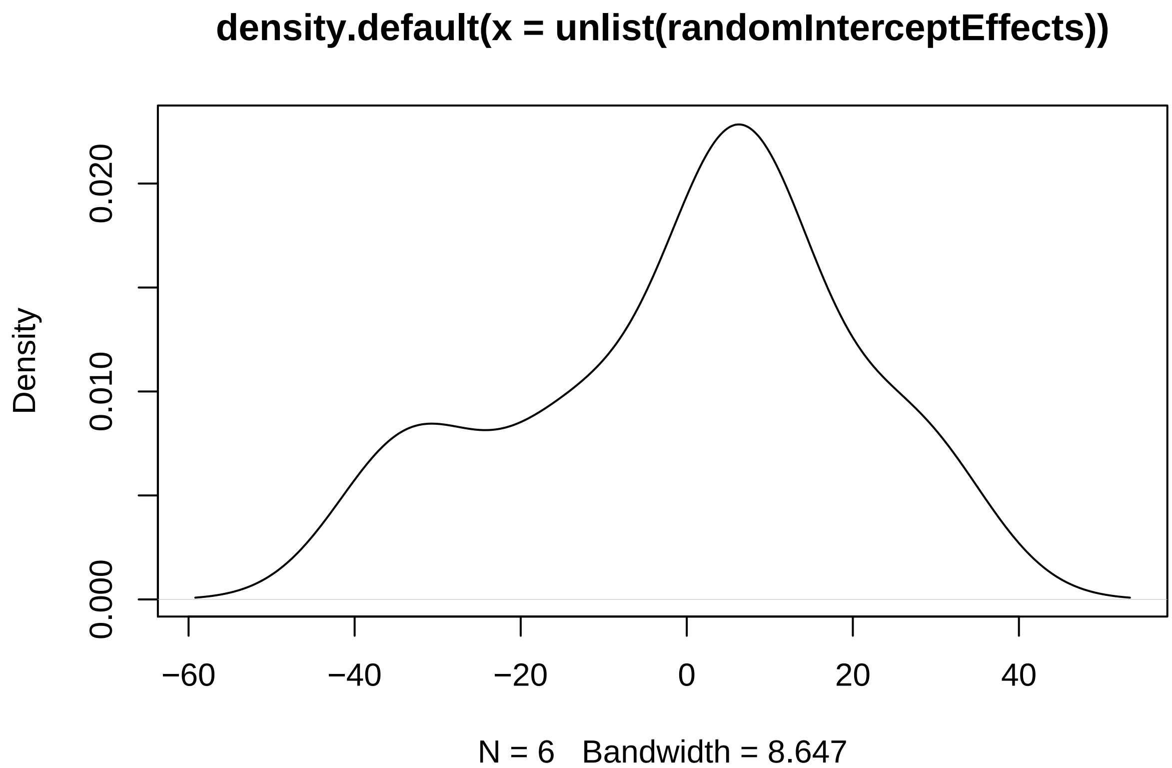
```r
randomInterceptVariance = 603.9 + 850.9
subjectVariance = 603.9/randomInterceptVariance
residualVariance = 850.9/randomInterceptVariance
subjectVariance
```

```
## [1] 0.4151086
```

```r
randomInterceptF = lmer(frequency ~ (1|subject) + gender + attitude, data = data, REML = FALSE)
randomInterceptEffects = ranef(randomIntercept)$subject
ranef(randomIntercept)
```

```
## $subject
##    (Intercept)
## F1  -13.582929
## F2   10.175840
## F3    3.407089
## M3   27.436373
## M4    5.818871
## M7  -33.255244
##
## with conditional variances for "subject"
```

```r
plot(density(unlist(randomInterceptEffects)))
```




The fixed effects of the model are about as expected. The baseline fequency for informal female participants have a pitch around 256.7 hz. Males tend to have a lower pitch compared to females, and speaking in a polite voice also tends to lower the pitch of the individual's voice. However, there is an effect on the frequency by subject which indicates that there is some variation in frequency by subject.

Testing the residuals for the random effects, variance by subject accounted for 41.5% of the total variance in the model, which is substantial and indicates that subject plays a significant role in determining pitch. There were 2 males that had more extreme variation that the others, but I think this signals it is proper to try and account for the subject as a random effect.

The KDE for the plot indicates that the random effects are not normally distributed.

### Random Intercept + Slope Model

```r
summary(randomSlope)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: frequency ~ (attitude | subject) + gender + attitude
##    Data: data
##
## REML criterion at convergence: 786.7
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.3484 -0.5488 -0.2010  0.4838  3.2156
##
## Random effects:
```

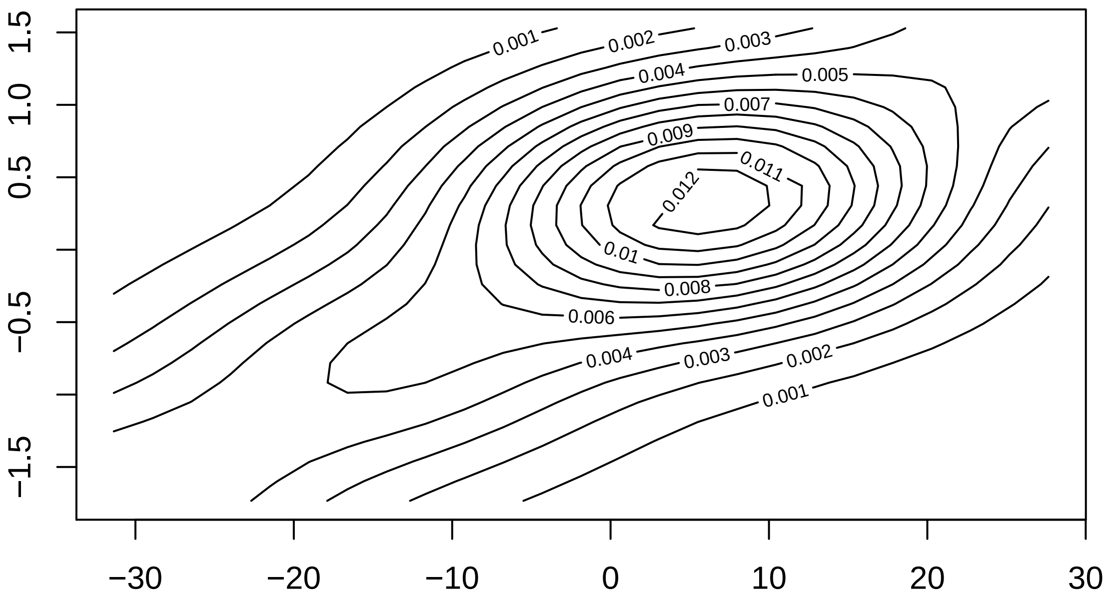
```
##  Groups   Name        Variance Std.Dev. Corr
##  subject  (Intercept) 572.596  23.929
##           attitudepol   1.749   1.322   1.00
##  Residual             850.548  29.164
## Number of obs: 83, groups:  subject, 6
##
## Fixed effects:
##             Estimate Std. Error       df t value Pr(>|t|)
## (Intercept)  257.749     15.031    4.116  17.148 5.53e-05 ***
## genderM     -110.322     21.001    4.009  -5.253  0.00624 **
## attitudepol  -19.402      6.428   59.004  -3.018  0.00375 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) gendrM
## genderM     -0.699
## attitudepol -0.157  0.003
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see ?isSingular
```

```r
randomSlopeF =  lmer(frequency ~ (attitude|subject) + gender + attitude, data = data, REML=FALSE)
```

```
## boundary (singular) fit: see ?isSingular
```

```r
randomSlopeSubjectEffect = ranef(randomSlope)$subject

contour(kde2d(randomSlopeSubjectEffect[,1], randomSlopeSubjectEffect[,2]))
```




```
varianceCorrelation = VarCorr(randomSlope) %>% as_data_frame() %>% mutate(icc=vcov/sum(vcov))
```

```
## Warning: `as_data_frame()` is deprecated as of tibble 2.0.0.
## Please use `as_tibble()` instead.
## The signature and semantics have changed, see `?as_tibble`.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_warnings()` to see where this warning was generated.
```

```
VPC = data.frame(varianceCorrelation$grp, varianceCorrelation$icc)
VPC
```

```
##   varianceCorrelation.grp varianceCorrelation.icc
## 1                 subject             0.393121313
## 2                 subject             0.001200757
## 3                 subject             0.021726557
## 4                Residual             0.583951373
```

The fixed effects for this model were slightly different from the other two models. THe base line pitch was 1hz lower, and the effects of being a male were more pronounced, but the polite attitude effect did not change.

The random effects for each indidvual were still fairly pronounced with the same two male participants as found in model 1.

For the slop effect, the VPC for the subject is 39.3%, .1%, and 2% for each attitude, with 58% of the variance being explained by the residuals. This is inline with the ICC where the subject accounted for 41.5% of the Variance overall, in this case the subject still takes up ~40% dependent on the attitude they are asked to used.

Based on the 2d KDE, the random effects are nearly normally distributed, though as they move away from the centers, they become less and less normal based on the warping of the contours towards the extremes.

This is consistent with the near normality of the other models' KDEs.

### Multiple Random Effects Model

```r
summary(multipleRandom)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: frequency ~ (1 | subject) + gender + attitude + (1 | scenario)
##    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
```

```r
multipleRandomVariance = 219.5 + 615.6 + 645.9
multipleScenarioVariance =219.5/multipleRandomVariance
multipleSubjectVariance = 615.6/multipleRandomVariance
multipleScenarioVariance
```

```
## [1] 0.1482107
```

```r
multipleSubjectVariance
```

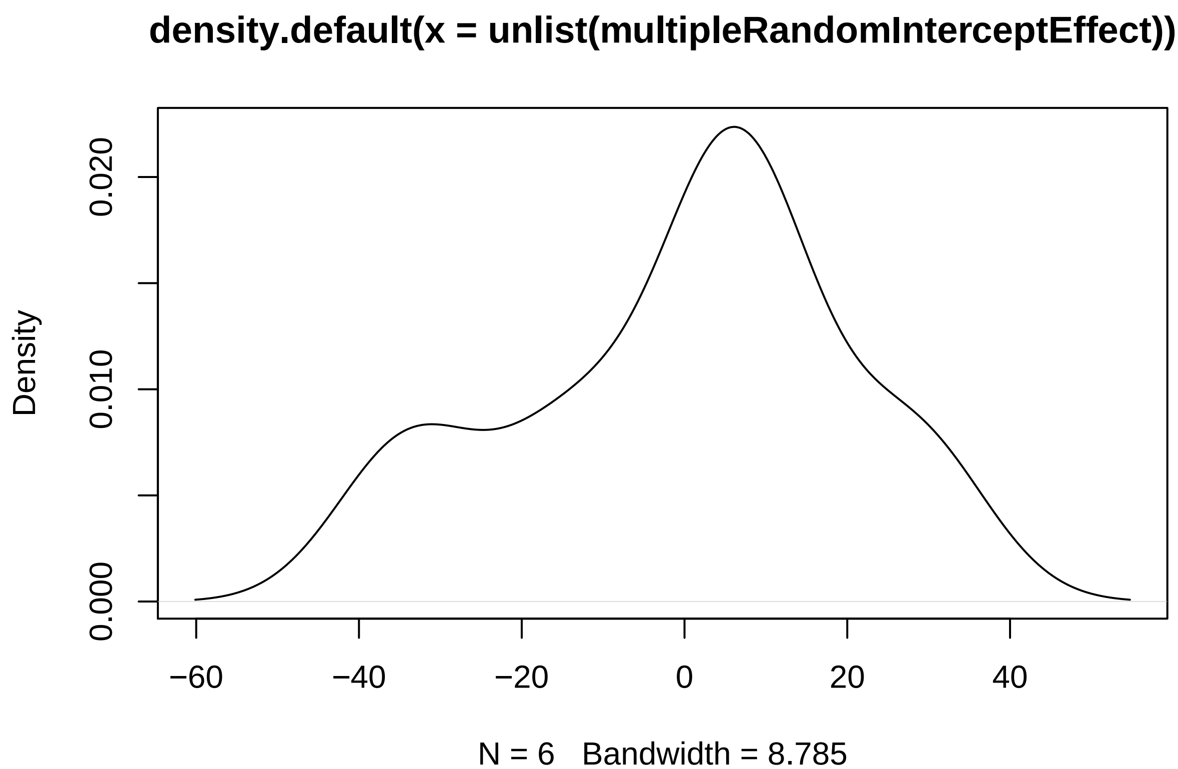
```
## [1] 0.4156651
```

```r
multipleRandomF = lmer(frequency ~ (1|subject) + gender + attitude + (1|scenario), data = data, REML = 
multipleRandomInterceptEffect = ranef(multipleRandom)$subject
multipleRandomInterceptEffectScenario = ranef(multipleRandom)$scenario
ranef(multipleRandom)
```

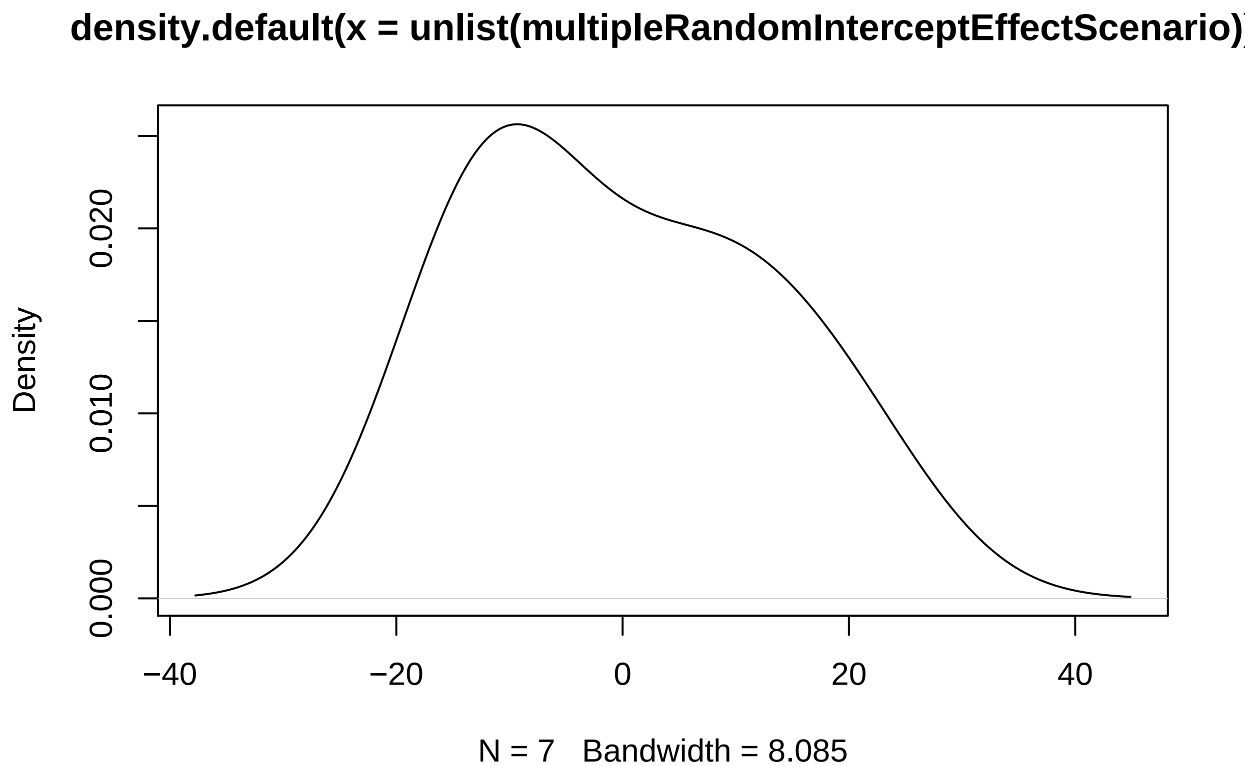
```
## $scenario
##   (Intercept)
## 1  -13.506509
```

```
## 2    6.582894
## 3   11.407831
## 4   20.629415
## 5   -1.936030
## 6  -12.173903
## 7  -11.003698
##
## $subject
##    (Intercept)
## F1  -13.907651
## F2   10.419110
## F3    3.488541
## M3   28.382033
## M4    5.378484
## M7  -33.760517
##
## with conditional variances for "scenario" "subject"
```

```r
plot(density(unlist(multipleRandomInterceptEffect)))
```




```r
plot(density(unlist(multipleRandomInterceptEffectScenario)))
```




The fixed effects of the model mimic what we saw in model 1. The baseline fequency for informal female participants have a pitch around 256.8 hz. Males tend to have a lower pitch compared to females, and speaking in a polite voice also tends to lower the pitch of the individual's voice.

However, there is more variation in the random efffects in this case. There is still 41.5% variance for the Subject, and an additional 14.8% of the variance is explained by the scenario that the subject was in, suggesting that subject is still a significant random effect when it comes to determining pitch, and also suggesting that scenario plays an important role.

The two male subjects still had more extreme values compared to the others, and secnario 4 seemed to stand out as well, suggesting it is correct to treat scenario as a separate random effect.

The KDE for the plot indicates that the random effects are more closely normally distributed than the other models, though they still appear to be bimodal for the subject.

The scenario effect KDE is close to normal but still skewe.

### AIC Comparison

```
AIC(randomInterceptF, randomSlopeF, multipleRandomF)
```

```
##                  df      AIC
## randomInterceptF  5 816.3376
## randomSlopeF      7 820.3094
## multipleRandomF   6 807.1015
```

```
anova(randomInterceptF, randomSlopeF)
```

```
## Data: data
```

```
## Models:
## randomInterceptF: frequency ~ (1 | subject) + gender + attitude
## randomSlopeF: frequency ~ (attitude | subject) + gender + attitude
##                   npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## randomInterceptF    5 816.34 828.43 -403.17   806.34
## randomSlopeF        7 820.31 837.24 -403.15   806.31 0.0282  2      0.986
```

```r
anova(randomSlopeF, multipleRandomF)
```

```
## Data: data
## Models:
## multipleRandomF: frequency ~ (1 | subject) + gender + attitude + (1 | scenario)
## randomSlopeF: frequency ~ (attitude | subject) + gender + attitude
##                  npar    AIC    BIC  logLik deviance Chisq Df Pr(>Chisq)
## multipleRandomF    6 807.10 821.61 -397.55   795.10
## randomSlopeF       7 820.31 837.24 -403.15   806.31     0  1          1
```

```r
anova(randomInterceptF, multipleRandomF)
```

```
## Data: data
## Models:
## randomInterceptF: frequency ~ (1 | subject) + gender + attitude
## multipleRandomF: frequency ~ (1 | subject) + gender + attitude + (1 | scenario)
##                   npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## randomInterceptF    5 816.34 828.43 -403.17   806.34
## multipleRandomF     6 807.10 821.61 -397.55   795.10 11.236  1  0.0008022 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Based on the AIC rating, the bvest model was model 3 due to the lowest AIC. the Anova comparison also shows that the multipleRAndomEffects is one of the better models, as it is the only model significantly different from the other models, and contains the least overall Deviance.