# A Tutorial for Linear Models and Linear Mixed Effects Models in R

## Experimental Phonetics

Ryu, Hyuksu

Naver Clova

October 11, 2017

# T.O.C

# Outline

# Before beginning

Source of this tutorial

- Bodo Winter hompage
  `http://www.bodowinter.com/tutorials.html`
- Linear models and linear mixed effects models in R
  `http://arxiv.org/pdf/1308.5499.pdf`

Citation

- Winter, B. (2013). Linear models and linear mixed effects models in R with linguistic applications. arXiv:1308.5499.

# Instruction

What we are dealing with?

1. Linear model
2. Linear mixed model $\leftarrow$

# Outline

# Fixed and random effects

In tutorial 1,

$$pitch \sim age + \epsilon$$

- age - fixed effect
  - systematic part
- $\epsilon$ - random factor
  - stochastic or probabilistic part
  - to represent the deviations from our predictions
  - "random" factors that we cannot control experimentally
- we will unpack this "$\epsilon$"

# Model Description

In this tutorial,

- interested in
  - relationship b/w pitch and politeness
  - relationship b/w pitch and sex

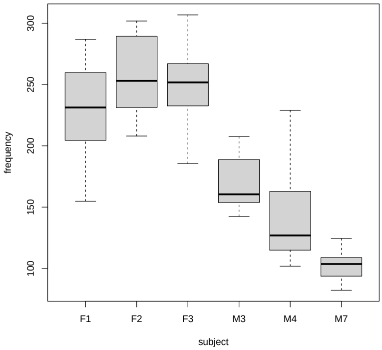
$$pitch \sim politeness + sex + \epsilon$$

- politeness
  - categorical factor with two levels - formal/informal
- **multiple** measures per subject
  - violation of independence assumption
- Every person has different voice pitch
  - rendering these different responses inter-dependent rather than independent.

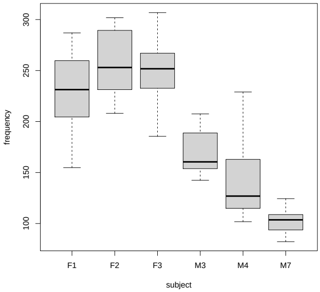
# Random effect

Dealing with the situation, Add **random effect** for subject

- to resolve this non-independence
  - by assuming a different "baseline" pithc value for each subject
  - for example, Subject 1 - 233 Hz (mean), and Subject 2 - 210 Hz



# Model description




From the box plot,

- pitch: male < female
- BUT, lots of individual variation

We can model these individual difference

- assuming different **random intercepts** for each subject

# Model description

Why “mixed” model?

- in addition to fixed effects,
- we add one or more random effects
  - part of $\epsilon$ that we cannot control for
  - in this case, a random effect for “subject”
  - this characterizes idiosyncratic variation that is due to individual differences

Updated formula

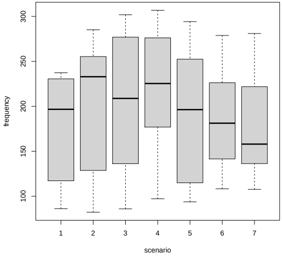
$$pitch \sim politeness + sex + (1|subject) + \epsilon$$

# Model description

One more random thing

- item!
  - 7 different scenarios (items)
  - examples
    - asking for a favor (polite/informal)
    - excusing for coming too late (polite/informal)
- also expect by-item variation
  - For example,
  - "excusing for coming too late" have overall higher pitch, regardless of the influence of politeness
  - ∵ it's more embarrassing than asking for a favor
  -

# Model description




Variation b/w items

- not as big as the variation b/w subjects
- but still noticeable, and better account for them in model

Adding an additional random effect

$$pitch \sim politeness + sex + (1|subject) + (1|item) + \epsilon$$

# Mixed model

The upshot is that

- mixed model give you much more flexibility than traditional model
- take the full data into account
- consider by-item variation and by-subject variation both in a single model

# Outline

# Pre-requisite

For a start,

- need to install the R package *lme4*

```
install.packages("lme4") # Linear Mixed Effect Model
```

- after installation, laod the *lme4* packages into R

```
library(lme4)
```

- a function we use for the mixed model
  - `lmer()`
  - the mixed model equivalent of the funtion `lm()` in linear model

# Loading data

## Loading data

```
politeness=read.csv("http://www.bodowinter.com/tutorial/politeness_data.csv")
```

```
summary(politeness)

##  subject gender    scenario attitude   frequency
##  F1:14   F:42   Min.   :1   inf:42   Min.   : 82.2
##  F2:14   M:42   1st Qu.:2   pol:42   1st Qu.:131.6
##  F3:14          Median :4            Median :203.9
##  M3:14          Mean   :4            Mean   :193.6
##  M4:14          3rd Qu.:6            3rd Qu.:248.6
##  M7:14          Max.   :7            Max.   :306.8
##                                      NA's   :1
```

- `scenario` should be factor
- should remove NA

# Loading data

- Factorize

```
politeness$scenario = factor(politeness$scenario)
```

- Remove NA

```
(idx = which(is.na(politeness$frequency)))

## [1] 39

politeness = politeness[-idx,]
```

- Summary

```
summary(politeness)

##  subject gender scenario attitude   frequency
##  F1:14   F:42   1:12     inf:42   Min.   : 82.2
##  F2:14   M:41   2:12     pol:41   1st Qu.:131.6
##  F3:14          3:12              Median :203.9
##  M3:14          4:12              Mean   :193.6
##  M4:13          5:12              3rd Qu.:248.6
##  M7:14          6:11              Max.   :306.8
##                 7:12
```

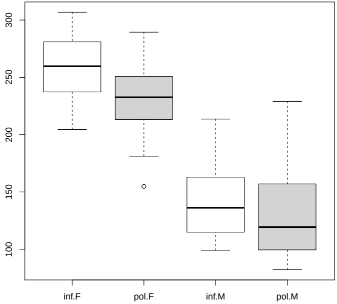
# Data structure

Variables

- Dependent measure
  - Frequency - Hz
- Fixed
  - Gender - M/F
  - Attitude - inf/pol
- Random
  - Subject
  - Scenario

# Data structure

Relationship b/w politeness & pitch




- polite < informal
- But more overlap b/w two politeness categories for males than for females

# Building mixed model

How to build mixed model?

```
politeness.model = lmer(frequency ~ attitude + (1|subject) +
          (1|scenario), data=politeness, REML=FALSE)
```

Display the full result

```
summary(politeness.model)
```

# Building mixed model

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: frequency ~ attitude + (1 | subject) + (1 | scenario)
##    Data: politeness
##
##      AIC      BIC   logLik deviance df.resid
##    817.0    829.1   -403.5    807.0       78
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.2127 -0.5906 -0.0598  0.5675  3.4584
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  scenario (Intercept)  216.8   14.72
##  subject  (Intercept) 3367.7   58.03
##  Residual              637.0   25.24
## Number of obs: 83, groups:  scenario, 7; subject, 6
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  202.588     24.646   8.220
## attitudepol  -19.692      5.546  -3.551
##
```

# Output

General summary statistics

```
##       AIC       BIC    logLik  deviance  df.resid
##  817.0395  829.1337 -403.5198  807.0395   78.0000
```

- Akaike's Information Criterion
- log Likelihood
- etc

# Output

Random effects

| Groups | Name | Variance | Std. Dev. |
|---|---|---|---|
| scenario | (Intercept) | 219 | 14.80 |
| subject | (Intercept) | 4015 | 63.36 |
| Residual | | 646 | 25.42 |

- less variation in subject than scenario
- which is expected by boxplot
- "Residual"
  - stands for the variability that is not due to scenario or subject - our $\epsilon$
  - "random" variation from the predicted that are not due to subjects and items

# Output

Fixed effect

```
##               Estimate Std. Error   t value
## (Intercept) 202.58810  24.645978  8.219925
## attitudepol -19.69216   5.545759 -3.550851
```

- slope
  - same way of interpretation with linear model
  - informal $\rightarrow$ polite: -19.695Hz
- intercept
  - pitch at informal state
  - fall halfway b/w males and females in the boxplot
  - average of our data for the informal condition
  - this is why we need gender as additional fixed effect

# Outline

# Building mixed model 2

How to build mixed model?

```
politeness.model.2 = lmer(frequency ~ attitude + gender + (1|subject) + (1|scen
```

Display the full result

```
summary(politeness.model.2)
```

# Building mixed model 2

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: frequency ~ attitude + gender + (1 | subject) + (1 | scenario)
##    Data: politeness
##
## 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 t value
## (Intercept)  256.846     16.116  15.938
## attitudepol  -19.721      5.584  -3.532
## genderM     -108.516     21.013  -5.164
##
```

# Output

Random effect

| Groups | Name | Variance | Std.Dev. |
|---|---|---|---|
| scenario | (Intercept) | 219.5 | 14.81 |
| subject | (Intercept) | 615.6 | 24.81 |
| Residual | | 645.9 | 25.42 |

- the variation of "subject" dropped drastically
- By adding the effect of gender, we have shifted a considerable amount of the variance that was previously in the random effects component (difference b/w male and female individuals) to the fixed effects component

# Output

Fixed effects

```
##                Estimate Std. Error   t value
## (Intercept)   256.84627  16.115629 15.937712
## attitudepol   -19.72111   5.584028 -3.531699
## genderM      -108.51635  21.013326 -5.164168
```

- female $>$ male about 109 Hz
- intercept is much higher
  - represents the female category (for informal condition)
  - effect of attitude didn't change much

# Outline

# p-value

p-value

- p-value for mixed models are not as straightforward as they are for the linear model
- There are multiple approaches (still controversial)
- **Likelihood Ratio Test** as a means to attain p-value
  - the probability of seeing the data you collected given your model
  - we compared a full model (with the fixed effects in question)
  - against a reduced model without the effects in question
  - by the result, we conclude that a fixed effect is significant if the difference b/w the likelihood of these two models is significant

# p-value

Likelihood Ratio Test

- Note: should add the argument REML=FALSE
- construct null model

```
politeness.null = lmer(frequency ~ gender + (1|subject)
        + (1|scenario), data=politeness, REML=FALSE)
```

- construct full model

```
politeness.model = lmer(frequency ~ attitude + gender
        + (1|subject) + (1|scenario), data=politeness, REML=FALSE)
```

- perform the likelihood ratio test using the anova()

```
anova(politeness.null, politeness.model)
```

full model: frequency $\sim$ attitude $+$ gender
reduced model: frequency $\sim$ gender

# p-value

## Result

```
## Data: politeness
## Models:
## politeness.null: frequency ~ gender + (1 | subject) + (1 | scenario)
## politeness.model: frequency ~ attitude + gender + (1 | subject) + (1 | scena
##                  Df    AIC    BIC  logLik deviance  Chisq Chi Df
## politeness.null   5 816.72 828.81 -403.36   806.72
## politeness.model  6 807.10 821.61 -397.55   795.10 11.618      1
##                  Pr(>Chisq)
## politeness.null
## politeness.model  0.0006532 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

## Report

"... politeness affected pitch ($\chi^2(1)=11.62$, p=0.00065), lowering it by about 19.7 Hz $\pm$ 5.6 (standard errors)..."

# interaction

How to say regarding interaction

- If you have any an inter-dependence b/w two factors?
- that is interaction
- can test it the following way $\rightarrow$ likelihood test
  full model: frequency $\sim$ attitude $+$ gender
  reduced model: frequency $\sim$ attitude * gender
- compare the models using `anova()`
- If `anova()` is significant, attitude and gender are significantly inter-dependent on each other

# interaction

```
politeness.model = lmer(frequency ~ attitude + gender + (1+attitude|subject)
        + (1+attitude|scenario), data=politeness, REML=FALSE)
politeness.inter = lmer(frequency ~ attitude * gender + (1+attitude|subject)
        + (1+attitude|scenario), data=politeness, REML=FALSE)
anova(politeness.model, politeness.inter)

## Data: politeness
## Models:
## politeness.model: frequency ~ attitude + gender + (1 + attitude | subject) +
## politeness.model:     attitude | scenario)
## politeness.inter: frequency ~ attitude * gender + (1 + attitude | subject) +
## politeness.inter:     attitude | scenario)
##                  Df    AIC    BIC  logLik deviance  Chisq Chi Df
## politeness.model 10 814.90 839.09 -397.45   794.90
## politeness.inter 11 814.89 841.50 -396.45   792.89 2.0023      1
##                  Pr(>Chisq)
## politeness.model
## politeness.inter     0.1571
```

No significant interaction here

# Outline

# random slopes vs. random intercepts

the coefficients of the model by subject and by item

```
coef(politeness.model)
```

# random slopes vs. random intercepts

```
## $scenario
##   (Intercept) attitudepol   genderM
## 1    245.2603   -20.43832 -110.8021
## 2    263.3012   -15.94385 -110.8021
## 3    269.1432   -20.63361 -110.8021
## 4    276.8309   -16.30132 -110.8021
## 5    256.0579   -19.40575 -110.8021
## 6    246.8605   -21.94816 -110.8021
## 7    248.4702   -23.55752 -110.8021
##
## $subject
##    (Intercept) attitudepol   genderM
## F1    243.8053   -20.68245 -110.8021
## F2    266.7321   -19.17028 -110.8021
## F3    260.1484   -19.60452 -110.8021
## M3    285.6958   -17.91950 -110.8021
## M4    264.1982   -19.33741 -110.8021
## M7    227.3551   -21.76744 -110.8021
##
## attr(,"class")
## [1] "coef.mer"
```

# random slopes vs. random intercepts

Result

- each scenario and subject is assigned **a different intercept**
  - by-subject and by-item variability
- the fixed effects (attitude and gender) are **all the same** for all subjects and items
- **random intercept model**
  - baseline-differences in pitch
  - but, effect of politeness and gender is the same for all subjects and items
- is it a valid assumption?
  - the effect of politeness might be different for diffent items and subjects
- so, we need a **random slope model**

# Random slope model

What we changed

- only the random effects
- `(1+attitude|subject)`
    - this notation means differing responses to attitude
    - as well as differing baseline-levels of frequency (intercept)

Model

```
politeness.model = lmer(frequency ~ attitude + gender
                       + (1+attitude|subject) + (1+attitude|scenario),
                       data=politeness, REML=FALSE)
```

Coefficients of this updated model

```
coef(politeness.model)
```

# Random slope model

```
## $scenario
##   (Intercept) attitudepol   genderM
## 1    245.2603   -20.43832 -110.8021
## 2    263.3012   -15.94385 -110.8021
## 3    269.1432   -20.63361 -110.8021
## 4    276.8309   -16.30132 -110.8021
## 5    256.0579   -19.40575 -110.8021
## 6    246.8605   -21.94816 -110.8021
## 7    248.4702   -23.55752 -110.8021
##
## $subject
##    (Intercept) attitudepol   genderM
## F1    243.8053   -20.68245 -110.8021
## F2    266.7321   -19.17028 -110.8021
## F3    260.1484   -19.60452 -110.8021
## M3    285.6958   -17.91950 -110.8021
## M4    264.1982   -19.33741 -110.8021
## M7    227.3551   -21.76744 -110.8021
##
## attr(,"class")
## [1] "coef.mer"
```

# Random slope model

Interpretation

- the column for politeness is different for each subject and item
- but it is always negative and quite similar to each other
- means that despite individual variation, there is also consistency in how politeness affects the voice
  - for all of speakers, the voice tends to go down in polite speech,
  - but for some people it goes down slightly more so than for others
- coefficients of gender do not change, since we did not specify random slope for the factor

# Random slope model

Obtain p-value in random slope model

- Likelihood ratio test
- Comparison of full model and the reduced model

```
politeness.model = lmer(frequency ~ attitude + gender
                        + (1+attitude|subject)+(1+attitude|scenario),
                        data=politeness, REML=FALSE)
politeness.null = lmer(frequency ~ gender
                       + (1+attitude|subject) + (1+attitude|scenario),
                       data=politeness, REML = FALSE)
```

- need the same random effects structure (random slope model)

# Random slope model

Likelyhood ratio test

```
anova(politeness.null, politeness.model)

## Data: politeness
## Models:
## politeness.null: frequency ~ gender + (1 + attitude | subject) + (1 + attitu
## politeness.null:     scenario)
## politeness.model: frequency ~ attitude + gender + (1 + attitude | subject) +
## politeness.model:     attitude | scenario)
##                  Df    AIC    BIC  logLik deviance  Chisq Chi Df
## politeness.null   9 819.61 841.37 -400.80   801.61
## politeness.model 10 814.90 839.09 -397.45   794.90 6.7082      1
##                  Pr(>Chisq)
## politeness.null
## politeness.model   0.009597 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Siginificant result!

# Why random slope?

Questions?

- Which random slopes should I specify?
- Are random slopes necessry at all?

Answer!

- it makes lots of sense to include random slopes most of the time
- it is reasonable to expect that the effect of an experimental manipulation is not going to be the same for all items.
- several researches have shown that mixed models w/o random slopes have a relatively high Type I error rate
  - They tend to find a lot of significant results which are actually due to chance

# Why random slope?

In this model, why did we have random slope for only attitude, not gender?

- we were not interested in gender differences
- they are well worth controlling for
- therefore, we only modeled by-subject and by-item variability in how politeness affects pitch

# Outline

# Assumptions in LM

Assumptions in LM

1. Linearity
2. Absense of collinearity
3. Homoskedasticity
4. Normality of residuals
5. Absense of influential data points
6. Independence

Good news!

- Everything applies straightforwardly to mixed model

# Independence

Why did we move to mixed model?

- to resolve non-independencies in our data

Cautions

- still should be cautious in violation
- missing fixed or random effects cause independence violation
- Choose fixed and random effects carefully

# Influential data points

Infuential data points in LM

- dfbeta() function

But, in the mixed model,

- dfbeta() function does not work
- "leave-one-out" diagnosis by hand
- example script for leave-one-out diagnosis for attitude

```
all.res.attitude = numeric(nrow(politeness))
all.res.gender = numeric(nrow(politeness))

for(i in 1:nrow(politeness)){
  myfullmodel = lmer(frequency ~ attitude + gender +
                       (1+attitude|subject)+(1+attitude|scenario),
                   data=politeness[-i,], REML=FALSE)
  # 1- intercept, 2-attitude, 3-gender
  all.res.attitude[i] = fixef(myfullmodel)[2]
  all.res.gender[i] = fixef(myfullmodel)[3] #
}
```

# Outline

# Final notes

More about fixed and random effects

- random effect
  - Generally something that can be expected to have
  - a non-systematic, idiosyncratic, unpredictable, or "random" influence on data
  - often "subject" and "item"
  - the things you generally want to generalize over the idiosyncrasies of individual subjects and items
  - generally "sample" from the population of interests
  - far away from "exhausting the population" b/c many many more subjects or items that you could have tested
- fixed effect
  - are expected to have a systematic and predictable influence on your data
  - "exhaust the population of interest"
  - "exhaust the levels of a factor"
  - For example, gender has only two levels and, by two categories, exhausts the category gender

# Write-up

More things to consider

1. Reproducibility
2. Coefficients
3. Citation

# Write-up

Reproducibility

- need to describe the model to such an extent that people can reproduce the analysis
- ask yourself the question
- "Would I be able to recreate the analysis given the information that I provided?"
- If the answer is "yes", write-up is good

Coefficients

- report the actual coefficients/estimates
- not just whether an effect is significant or not
- also mention standard errors

# Write-up

Citation for R: citation()

```
##
## To cite R in publications use:
##
##   R Core Team (2017). R: A language and environment for
##   statistical computing. R Foundation for Statistical Computing,
##   Vienna, Austria. URL https://www.R-project.org/.
##
## A BibTeX entry for LaTeX users is
##
##   @Manual{,
##     title = {R: A Language and Environment for Statistical Computing},
##     author = {{R Core Team}},
##     organization = {R Foundation for Statistical Computing},
##     address = {Vienna, Austria},
##     year = {2017},
##     url = {https://www.R-project.org/},
##   }
##
## We have invested a lot of time and effort in creating R, please
## cite it when using it for data analysis. See also
## 'citation("pkgname")' for citing R packages.
```

# Write-up

Citation for package ‘lme4’

```
citation('lme4')

##
## To cite lme4 in publications use:
##
##   Douglas Bates, Martin Maechler, Ben Bolker, Steve Walker (2015).
##   Fitting Linear Mixed-Effects Models Using lme4. Journal of
##   Statistical Software, 67(1), 1-48. doi:10.18637/jss.v067.i01.
##
## A BibTeX entry for LaTeX users is
##
##   @Article{,
##     title = {Fitting Linear Mixed-Effects Models Using {lme4}},
##     author = {Douglas Bates and Martin M{\"a}chler and Ben Bolker and Steve
##     journal = {Journal of Statistical Software},
##     year = {2015},
##     volume = {67},
##     number = {1},
##     pages = {1--48},
##     doi = {10.18637/jss.v067.i01},
##   }
```

# Example Writing

### Example

We used R (R Core Team, 2014) and *lme4* (Bates, Maechler, Bolker & Walker, 2014) to perform a linear mixed effects analysis of the relationship between pitch and politeness. As fixed effects, we entered politeness and gender (without interaction term) into the model. As random effects, we had intercepts for subjects and items, as well as by-subject and by-item random slopes for the effect of politeness. Visual inspection of residual plots did not reveal any obvious deviations from homoscedasticity or normality. P-values were obtained by likelihood ratio tests of the full model with the effect in question against the model without the effect in question.