PANEL DATA ANALYSIS: LINEAR MIXED-EFFECT MODELS

Introduction and Motivation

Panel Data Econometrics vs. Mixed Models

- Pane data econometrics has its counterparts in the statistic literature on *mixed effect*, *hierarchical models*, or *models for longitudinal data*.
- In part they are just differences in terminologies used, but in part there are substantial distintion.
- For mixed models, $ has the long-standing **nlme** package (Pinheiro et al. 2007) and the more recent **lme4** package (Bates 2007).
- In this course, we will use the **lme4** package.

Introduction and Motivation

pitch

- This is a pretty common example: modeling \$ as a function of *age*

$$pitch_i = age_i + \epsilon_i$$

where $i = 1, 2, \ldots, n$

- We called "age" a fixed effect, and ϵ as "error term" to represent the deviations from predictions due to "random" factors that we are not able to control for experimentally.

- Note that the use of the term "fixed effect" here, which may be confusing to you.

- In the mixed model, we add one or more random effects. These random effects give structure to the error term ϵ, which in the linear regression model should not have any structure.

- Now, consider the study Winter & Grawunder, 2012. "The Phonetic Profile of Korean Formality", Journal of Phonetics. They studied the relationship between *pitch* and *politness*, and they add $ as an additional fixed effect. Using *R*, the model is

$$pitch \sim politness + \epsilon$$

where *politness* is entered as a binary variable with two levels.

- Add an additional fixed effect: *sex*

$$pitch \sim politness + sex + \epsilon$$

- Their study took multiple measures per subjects: each subject gave multiple polite responses and multiple informal responses.

- Multiple measures per subjects immediately would violate the independence assumption (in linear regression models),as multiple responses from the same subject cannot be regarded as independent from each other

- Importantly, every person has a slightly different voice pitch; thus, this is an individual-specific factor that affects all responses from the same subject, rendering these different responses inter-dependent rather than independent.
- We will model this individual-specific factor as a random effect for the subjects, allowing for a different "baseline" pitch value for each subject. Note that this type of situation is very common in many different applications in practice.

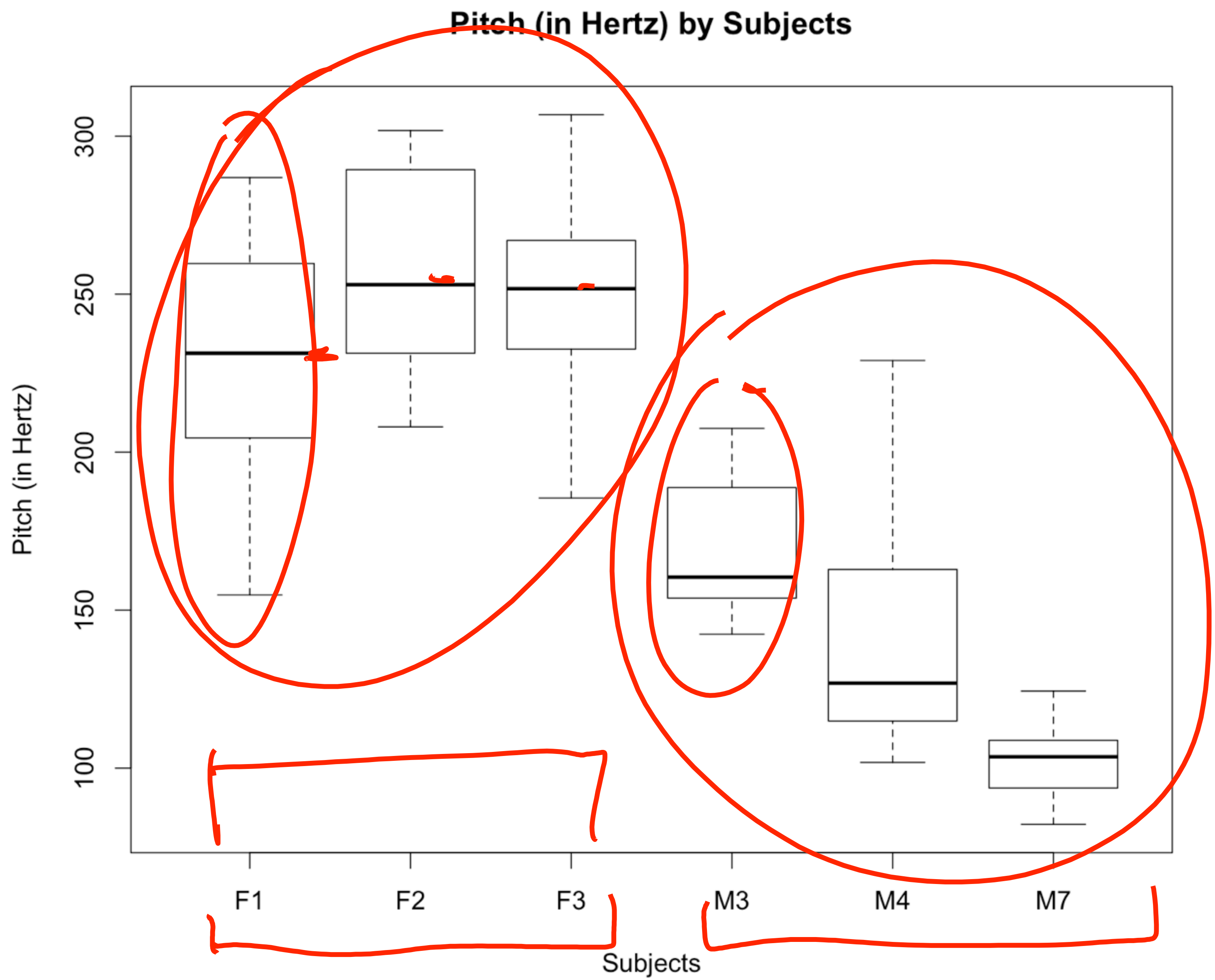
Let's look at the data:

```
> politeness= read.csv("http://www.bodowinter.com/tutorial/politeness_data.csv")
> str(politeness)
'data.frame':   84 obs. of  5 variables:
 $ subject  : Factor w/ 6 levels "F1","F2","F3",..: 1 1 1 1 1 1 1 1 1 1 ...
 $ gender   : Factor w/ 2 levels "F","M": 1 1 1 1 1 1 1 1 1 1 ...
 $ scenario : int  1 1 2 2 3 3 4 4 5 5 ...
 $ attitude : Factor w/ 2 levels "inf","pol": 2 1 2 1 2 1 2 1 2 1 ...
 $ frequency: num  213 204 285 260 204 ...
```

```
> table(politeness$subject)

F1 F2 F3 M3 M4 M7
14 14 14 14 14 14
```

```
> which(is.na(politeness$frequency))
[1] 39
```


Pitch (in Hertz) by Subjects
Pitch (in Hertz)
300
250
200
150
100
F1
F2
F3
M3
M4
M7
Subjects

- We can model these individual differences using different (random) intercepts for each subject: each subject has a different estimated intercept, and the mixed model can be used to estimate these intercepts.

- In the mixed model, we add one or more random effects to the fixed effects. These random effects give a structure to the error term ϵ.

- In our example, a random effect for "subject" is added, and this characterizes idiosyncratic variation that is due to individual differences.

- The updated model is

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

- This formula, in R notation, models "an intercept that's different for each subject" and "1" stands for the intercept.

- Think of this formula as telling the model that it should expect multiple responses per subject and the responses depend on each subject's baseline level.

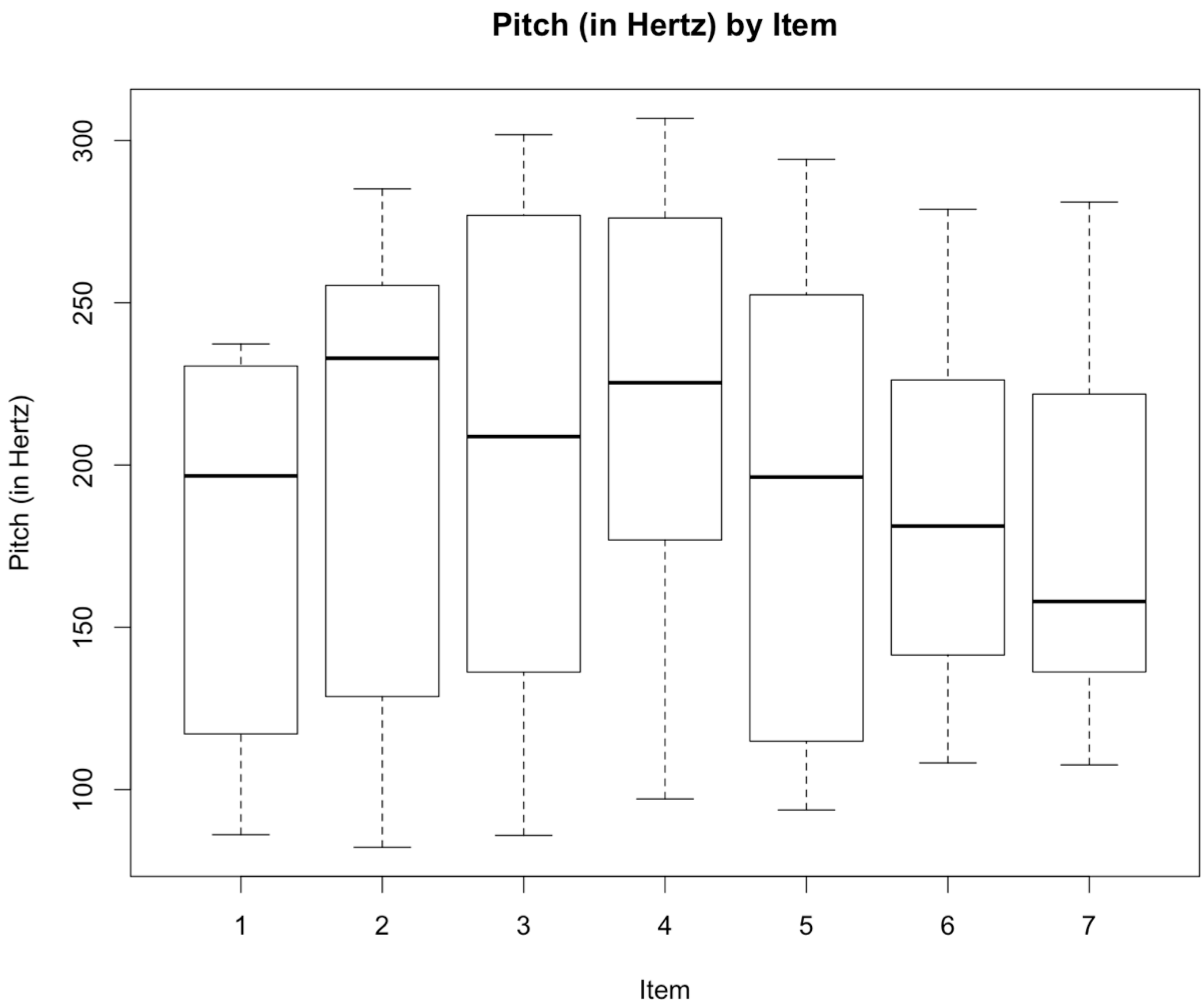
- Importantly, this framework resolves the non-independence that stems from having multiple responses by the same subject.

- In Winter and Grawunder's study, there were different items. One item, for example, was an "asking for a favor" scenario. Subjects had to imagine asking a professor for a favor (polite condition), or asking a peer for a favor (informal condition). Another item was an "excusing for coming too late" scenario, which was similarly divided between polite and informal. There were 7 different items in their study.

- Hence, we should also expect "*item-specific*" variation.

- Most importantly, the different responses to one item cannot be regarded as independent. If these interdependencies are not accounted for, the independence assumption would be violated.

By-Item Variation

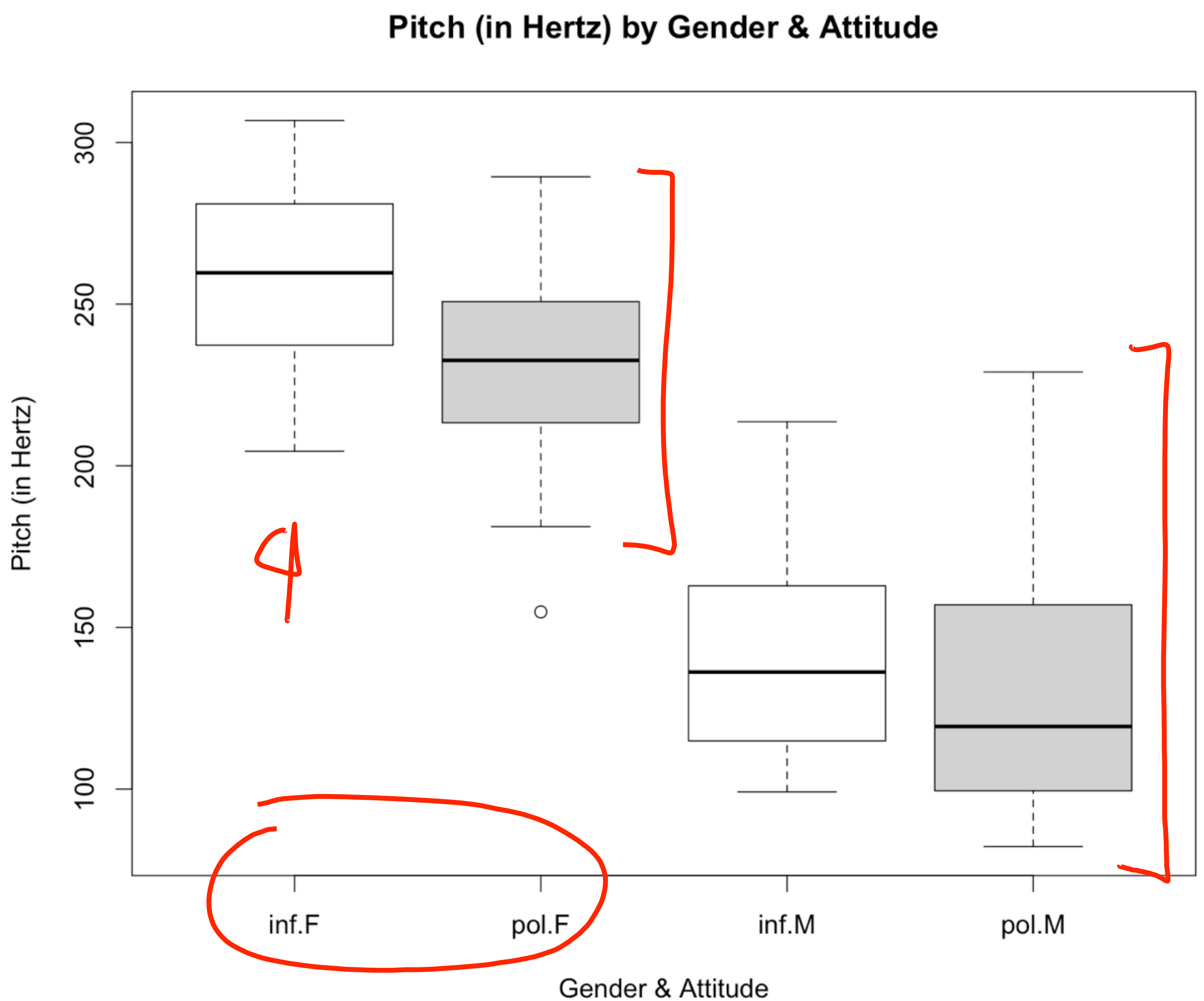


- The model now becomes

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

- Note that in addition to different intercepts for different subjects, the model includes different intercepts for different items. As such, non-independencies are accounted for - the model knows that there are multiple responses per subject and per item, and there are by-subject and by-item variation in overall pitch levels.

- Let's look at the relationship between politeness and pitch using a boxplot:


Pitch (in Hertz) by Gender & Attitude
Pitch (in Hertz)
300
250
200
150
100
inf.F
pol.F
inf.M
pol.M
Gender & Attitude

- Estimate the model using functions in R's lme4 package:

```
politeness.lmm = lmer(frequency ~ attitude + (1|subject) + (1|scenario),
data=politeness)
summary(politeness.lmm)
```

```
Linear mixed model fit by REML ['lmerMod']
Formula: frequency ~ attitude + (1 | subject) + (1 | scenario)
   Data: politeness

REML criterion at convergence: 793.5

Scaled residuals:
    Min      1Q  Median      3Q     Max
-2.2006 -0.5817 -0.0639  0.5625  3.4385

Random effects:
 Groups   Name        Variance Std.Dev.
 scenario (Intercept)  219     14.80
 subject  (Intercept) 4015     63.36
 Residual              646     25.42
Number of obs: 83, groups:  scenario, 7; subject, 6

Fixed effects:
            Estimate Std. Error t value
(Intercept)  202.588     26.754   7.572
attitudepol  -19.695      5.585  -3.527

Correlation of Fixed Effects:
            (Intr)
attitudepol -0.103
```

- One of the biggest differences between this model output and those we have seen earlier in this course is the inclusion of the Random effects:

- Have a look at the column standard deviation under random effects. This measures of how much variability in the dependent measure there is due to scenarios and subjects (our two random effects).

- The scenario ("item") has much less variability than subject, which we have already seen in the boxplots from above. THere is more idiosyncratic differences between subjects than between items, as expected.

- "Residual"stands for the variability that's not due to either scenario or subject. This corresponds to ϵ, the "random" deviations from the predicted values that are not due to subjects and items.

- It reflects the fact that each and every utterance has some factors that affect pitch that are outside of the purview of our experiment.

- The fixed effects output mirrors those in linear regression models

- The coefficient "*attitudepol*" is the slope for the categorical effect of politeness. -19.695 means that the pitch on average goes down by -19.695 Hz when going from "informal" to "polite". That is, pitch is lower in polite speech than in informal speech, by about 20 Hz.

- The model intercept 202.588 Hz is the average of our data for the informal condition, without distinguishing the gender difference. As we didn't inform our model that there are two sexes in our dataset, the intercept is particularly off, in between the voice pitch of males and females.

- Let's add gender to the model and re-estimate it again.

```
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

Correlation of Fixed Effects:
            (Intr) atttdp
attitudepol -0.173
genderM     -0.652  0.004
```

- We added "gender" as a fixed effect because the relationship between sex and pitch is systematic and predictable
- Compared to the last model without the gender fixed effect, the variation associated with the random effect "subject" dropped considerably.
- This is because the variation that's due to gender was confounded with the variation that's due to subject.
- Under the Fixed effects, we see that males and females differ by about 109 Hz. The intercept is now much higher (256.846 Hz), as it now represents the female category under the informal condition.
- The coefficient for the effect of attitude didn't change much.

- Importantly, *p-values* for mixed models are not as straightforward as they are for the linear regression model.

- We can use the **Likelihood Ratio Test** as a means to attain p-values. To implement it in *R*, we will have to re-run the model twice, but setting the argument *REML* to FALSE.

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

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

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

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

- We would report the result the following way: "… politeness affected pitch ($\chi^2(1) = 11.62, p = 0.00065$), lowering it by about $19.7Hz \pm 5.6$ (standard errors) …"

- Note that we kept the predictor "gender"" in the model. The only change between the full model and the null model that we compared in the likelihood ratio test was the factor of interest, *politeness*. In this particular test, think of "gender" as a control variable and of "attitude" as the test variable.

- **Interaction:** What happens if we have an interaction term? Suppose we predicted "attitude" to have an effect on pitch that is somehow modulated through "gender".

- For example, it could be that speaking politely versus informally has the opposite effect for men and women. Or, it could be that women show a difference and men don't (or vice versa)

full model:

$$frequency \sim attitude * gender$$

reduced model:

$$frequency \sim attitude + gender$$

- Comparison of the above models in a likelihood ratio test using the *anova*() function produces a p-value that gives the significance of the interaction.
- If it is significant, attitude and gender are significantly inter-dependent on each other.

Random Intercept vs. Random Slope:

- Note that each scenario and each subject is assigned a different intercept. That's what we would expect, given that we've told the model with "(1|subject)" and "(1|scenario)" to take by-subject and by-item variability into account.

- On the other hand, the fixed effects (attitude and gender) are all the same for all subjects and items. This model is called a **random intercept model**.

- In this model, we account for baseline-differences in pitch, but we assume that whatever the effect of politeness is, it's going to be the same for all subjects and items.

```
> coef(politeness.lmm2)
$scenario
  (Intercept) attitudepol    genderM
1    243.3398   -19.72111 -108.5163
2    263.4292   -19.72111 -108.5163
3    268.2541   -19.72111 -108.5163
4    277.4757   -19.72111 -108.5163
5    254.9102   -19.72111 -108.5163
6    244.6724   -19.72111 -108.5163
7    245.8426   -19.72111 -108.5163

$subject
   (Intercept) attitudepol    genderM
F1    242.9386   -19.72111 -108.5163
F2    267.2654   -19.72111 -108.5163
F3    260.3348   -19.72111 -108.5163
M3    285.2283   -19.72111 -108.5163
M4    262.2248   -19.72111 -108.5163
M7    223.0857   -19.72111 -108.5163

attr(,"class")
[1] "coef.mer"
```

- However, it may not be a valid assumption. For instance, it is possible that some items would elicit more or less politeness. That is, the effect of politeness might be different for different items. Likewise, the effect of politeness might be different for different subjects.

- In this case, we need a **random slope model**, where subjects and items are not only allowed to have differing intercepts, but different slopes for the effect of politeness.

```
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: frequency ~ attitude + gender + (1 + attitude | subject) + (1 +
    attitude | scenario)
   Data: politeness

     AIC      BIC   logLik deviance df.resid
   814.9    839.1   -397.4    794.9       73

Scaled residuals:
    Min      1Q  Median      3Q     Max
-2.1947 -0.6691 -0.0789  0.5256  3.4252

Random effects:
 Groups   Name        Variance Std.Dev. Corr
 scenario (Intercept) 182.082  13.494
          attitudepol  31.262   5.591   0.22
 subject  (Intercept) 392.474  19.811
          attitudepol   1.707   1.307   1.00
 Residual             627.880  25.058
Number of obs: 83, groups:  scenario, 7; subject, 6

Fixed effects:
            Estimate Std. Error t value
(Intercept)  257.989     13.529  19.069
attitudepol  -19.747      5.922  -3.334
genderM     -110.802     17.512  -6.327

Correlation of Fixed Effects:
            (Intr) atttdp
attitudepol -0.105
genderM     -0.647  0.003
```

- Note that the only thing being changed is the random effects: “$(1 + attitude|subject)$” means that you inform the model to expect differing baseline-levels of frequency (the intercept, represented by 1) as well as differing responses to the main factor in question, which is “attitude” in this case.

- Let’s look at the coefficients of this updated model:

```
$scenario
  (Intercept) attitudepol   genderM
1    245.2603   -20.43832 -110.8021
2    263.3012   -15.94386 -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   -17.91951 -110.8021
M4    264.1982   -19.33741 -110.8021
M7    227.3551   -21.76744 -110.8021
```

- Now, with the by-subject and by-item coefficients, the effect of politeness ("attitudepol") is different for each subject and item.
- **Interpretation:** Note the coefficients are always negative and quite similar to each other, meaning that despite individual variation, there is also consistency in how politeness affects the voice: for all of our speakers, the voice tends to go down when speaking politely, but for some people it goes down slightly more so than for others.
- The coefficients for gender have not changed, as we did not specify random slopes for the by-subject or by-item effect of gender.
- Let's obtain the *p-value* using likelihood ratio test:
- Note that the null model needs to have the same random effects structure. If the full model is a random slope model, then the null model also needs to be a random slope model.

```
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: frequency ~ attitude + gender + (1 + attitude | subject) + (1 +
    attitude | scenario)
   Data: politeness

     AIC      BIC   logLik deviance df.resid
   814.9    839.1   -397.4    794.9       73

Scaled residuals:
    Min      1Q  Median      3Q     Max
-2.1947 -0.6691 -0.0789  0.5256  3.4252

Random effects:
 Groups   Name        Variance Std.Dev. Corr
 scenario (Intercept) 182.082  13.494
          attitudepol  31.262   5.591   0.22
 subject  (Intercept) 392.474  19.811
          attitudepol   1.707   1.307   1.00
 Residual             627.880  25.058
Number of obs: 83, groups:  scenario, 7; subject, 6

Fixed effects:
            Estimate Std. Error t value
(Intercept)  257.989     13.529  19.069
attitudepol  -19.747      5.922  -3.334
genderM     -110.802     17.512  -6.327

Correlation of Fixed Effects:
            (Intr) atttdp
attitudepol -0.105
genderM     -0.647  0.003
```

```
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: frequency ~ gender + (1 + attitude | subject) + (1 + attitude |      scenario)
   Data: politeness

     AIC      BIC   logLik deviance df.resid
   819.6    841.4   -400.8    801.6       74

Scaled residuals:
     Min       1Q   Median       3Q      Max
-2.09488 -0.64641 -0.08678  0.60655  3.00533

Random effects:
 Groups   Name        Variance Std.Dev. Corr
 scenario (Intercept) 231.845  15.226
          attitudepol 410.122  20.251   -0.40
 subject  (Intercept) 378.499  19.455
          attitudepol   5.439   2.332   1.00
 Residual             628.654  25.073
Number of obs: 83, groups:  scenario, 7; subject, 6

Fixed effects:
            Estimate Std. Error t value
(Intercept)   253.37      13.44  18.856
genderM      -112.49      17.47  -6.439

Correlation of Fixed Effects:
        (Intr)
genderM -0.650
```

```
> anova(politeness.null,politeness.lmm3)
Data: politeness
Models:
politeness.null: frequency ~ gender + (1 + attitude | subject) + (1 + attitude |
politeness.null:     scenario)
politeness.lmm3: frequency ~ attitude + gender + (1 + attitude | subject) + (1 +
politeness.lmm3:     attitude | scenario)
                Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
politeness.null  9 819.61 841.37 -400.80   801.61
politeness.lmm3 10 814.90 839.09 -397.45   794.90 6.7082      1   0.009597 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

- What kind of random slope structure, if at all, should be imposed?
- It turns out that random slopes are very useful in practice. In experimental settings, experimental subjects often differ with how they react to an experimental manipulation, and the effect of an experimental manipulation differs for different items.
- **Assumptions:** We will study them in details in the next section. For now, we just want to highlight the importance of the *independence* assumption.
- Mixed models can still violate independence if important fixed or random effects are excluded in the model.
- In fact, missing some of the important variables may also lead to omitted variable bias, if these variables are correlated with the explanatory variables included in the model.
- For example, if we analyzed our data with a model that excluded the "subject-specific" random effect, then the model would not "know" that there are multiple responses per subject. This leads to a violation of the independence assumption. The punchline is that choose fixed effects and random effects carefully, and always try to resolve non-independencies when using data with repeated measurements.

Theory

- For hierarchical data with a single level of grouping, we can formulate the classical LMM at a given level of a grouping factor as follows

$$\mathbf{y_i} = \mathbf{X_i}\beta + \mathbf{Z_i}\mathbf{b_i} + \epsilon_\mathbf{i}$$

where $\mathbf{y_i}, \mathbf{X_i}, \beta$, and are the vectors of continuous responses, the design matrix, and the vector of residual errors for group $i, i = 1, 2, \ldots, n$, and $\mathbf{Z_i}$ and $\mathbf{b_i}$ are the matrix of covariates and the corresponding vector of random effects:

$$\mathbf{Z_i} \equiv \begin{pmatrix} \boxed{\begin{matrix} z_{i1}^{(1)} \\ \vdots \\ z_{in_i}^{(1)} \end{matrix}} & \begin{matrix} z_{i1}^{(2)} \\ \vdots \\ z_{in_i}^{(2)} \end{matrix} & \begin{matrix} \cdots \\ \ddots \\ \cdots \end{matrix} & \begin{matrix} z_{i1}^{(q)} \\ \vdots \\ z_{in_i}^{(q)} \end{matrix} \end{pmatrix}$$

$Z_i^{(k)}$

where each of the columns in $\mathbf{Z_i}$ can be more compactly represented as $\underline{\left(\mathbf{z}^{(}\mathbf{k})_\mathbf{i}\right)}$ and $\mathbf{b_i} \equiv (b_{i1}, \ldots, b_{iq})^T$

- Similar to the design matrix $\mathbf{X_i}$, the matrix $\mathbf{Z_i}$ contains known values of q covariates, with corresponding unobservable effects $\mathbf{b_i}$

Importantly, $\mathbf{b_i} \sim N_q(\mathbf{0}, \mathbf{D}), \epsilon_\mathbf{i} \sim N_{n_i}(\mathbf{0}, \mathbf{R_i}) with \mathbf{b_i} \perp \epsilon_\mathbf{i}$

- That is, the residual errors $\epsilon_\mathbf{i}$ for the same group are independent of the random effects $\mathbf{b_i}$.

- Also assume that vectors of random effects and residual errors for different groups are independent of each other: $b_i \perp \epsilon_{i'}$ for $i \neq i'$

$$\mathcal{D} = \sigma^2 \boldsymbol{D} \quad \text{and} \quad \mathcal{R}_i = \sigma^2 \boldsymbol{R}_i;$$

where σ^2 s an unknown scale parameter. In general, we will assume that $\mathbf{D}$ and $\mathbf{R_i}$ are positive-definite, unless stated otherwise

- In addition to the fixed-effects parameters $\mathbf{b}$ for the covariates used in constructing the design matrix $\mathbf{X_i}$, the proposed model includes two random components: the within-group residual errors $\$$ and the random effects $\mathbf{b_i}$ for the covariates included in the matrix $\mathbf{Z_i}$. The presence of fixed and random effects of known variables gives rise to the name of this class of models.

- Some authors, researchers, partitioners refer to the proposed framework as a *two-level* or *two-stage* model. Following authors such as Pinheiro and Bates (2000), we will refer it to as a single-level linear mixed model.

- The framework established so far can easily be extended to multilevel grouped data.

- For instance, a model for data with two levels of grouping, with observations grouped into N first-level groups (indexed by $i = 1, \ldots, N$), each with n_i second-level (sub-)groups (indexed by $j = 1, \ldots, n_i$) containing n_{ij} observations, can be written as

$$\mathbf{y}_{ij} = \boldsymbol{X}_{ij}\boldsymbol{\beta} + \mathbf{Z}_{1,ij}\mathbf{b}_i + \mathbf{Z}_{2,ij}\mathbf{b}_{ij} + \boldsymbol{\varepsilon}_{ij}$$

$$\mathbf{b}_i \sim \mathcal{N}_{q_1}(\mathbf{0}, \mathcal{D}_1), \quad \mathbf{b}_{ij} \sim \mathcal{N}_{q_2}(\mathbf{0}, \mathcal{D}_2), \quad \text{and} \quad \boldsymbol{\varepsilon}_{ij} \sim \mathcal{N}_{n_{ij}}(\mathbf{0}, \mathcal{R}_{ij})$$

where the random vectors $\mathbf{b}_i$, $\mathbf{b}_{ij}$, and $\boldsymbol{\varepsilon}_{ij}$ are independent of each other.

- In this setup, $\mathbf{b_i}$ are the random effects associated with the first-level groups, while $\mathbf{b_{ij}}$ are the random effects, independent of the first-level random effects, associated with the second-level groups. The design matrices $\mathbf{Z_{1,ij}}$ and $\mathbf{Z_{2,ij}}$ could be identical. Following *Pinheiro and Bates (2000)*, we refer this model as a *two-level LMM*.

- It follows that, for the classical LMMs, the conditional distribution, $f_{y|b}(\mathbf{y_i}|\mathbf{b_i})$, of $\mathbf{y_i}$ given $\mathbf{b_i}$ is multivariate normal, with the mean and variance defined as:

$$\mathrm{E}(\mathbf{y}_i|\mathbf{b}_i) \equiv \boldsymbol{\mu}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i$$

$$\mathrm{Var}(\mathbf{y}_i|\mathbf{b}_i) = \sigma^2\boldsymbol{R}_i,$$

with $\boldsymbol{\mu}_i \equiv (\mu_{i1}, \ldots, \mu_{i,n_i})'$

$$\mathrm{E}(y_{ij}|\mathbf{b}_i) \equiv \mu_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{b}_i;$$

where $\mathbf{x}_{ij} \equiv (x_{ij}^{(1)}, \ldots, x_{ij}^{(p)})'$ and $\mathbf{z}_{ij} \equiv (z_{ij}^{(1)}, \ldots, z_{ij}^{(q)})'$ are column vectors

- Conditionally on the (unknown) values of the random effects $\mathbf{b_i}$, the mean value of the dependent-variable vector $\mathbf{y_i}$ is defined by a linear combination of the vectors of the $\mathbf{X}$- and $\mathbf{Z}$-covariates included, as columns, in the group-specific design matrices $\mathbf{X_i}$ and $\mathbf{Z_i}$, corresponding to the fixed effects $\$$ and random effects $\mathbf{b_i}$.

- Moreover, the conditional variance-covariance matrix of $\mathbf{y_i}$ is equal to the variance-covariance matrix of the residual errors $\epsilon_\mathbf{i}$

Maximum-Likelihood Estimation

$$\ell_{\text{Full}}(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\theta}) \equiv -\frac{N}{2}\log(\sigma^2) - \frac{1}{2}\sum_{i=1}^{N}\log[\det(\boldsymbol{V}_i)]$$

$$-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(\mathbf{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})' \boldsymbol{V}_i^{-1}(\mathbf{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})$$

$$\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}) \equiv \left(\sum_{i=1}^{N} \boldsymbol{X}_i' \boldsymbol{V}_i^{-1} \boldsymbol{X}_i\right)^{-1} \sum_{i=1}^{N} \boldsymbol{X}_i' \boldsymbol{V}_i^{-1} \mathbf{y}_i,$$

$$\widehat{\sigma}^2_{\text{ML}}(\boldsymbol{\theta}) \equiv \sum_{i=1}^{N} \mathbf{r}_i' \boldsymbol{V}_i^{-1} \mathbf{r}_i / n,$$

where $\mathbf{r}_i \equiv \mathbf{r}_i(\boldsymbol{\theta}) = \mathbf{y}_i - \boldsymbol{X}_i \widehat{\boldsymbol{\beta}}(\boldsymbol{\theta})$.

$$\widehat{\sigma}^2_{\text{REML}}(\boldsymbol{\theta}) \equiv \sum_{i=1}^{N} \mathbf{r}_i' V_i^{-1} \mathbf{r}_i / (n-p)$$

- This leads to a log-profile-restricted-likelihood function, which only depends on θ:

$$\ell^*_{\text{REML}}(\boldsymbol{\theta}) \equiv -\frac{n-p}{2} \log\left(\sum_{i=1}^{N} \mathbf{r}_i' \mathbf{r}_i\right) - \frac{1}{2} \sum_{i=1}^{N} \log[\det(\boldsymbol{V}_i)]$$

$$-\frac{1}{2} \log\left[\det\left(\sum_{i=1}^{N} \boldsymbol{X}_i' \boldsymbol{V}_i^{-1} \boldsymbol{X}_i\right)\right].$$

To sum up, we have:

$$\mathcal{Y} \sim \mathcal{N}(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{o}, \sigma^2 \boldsymbol{W}^{-1})$$

$$(\mathcal{Y}|\mathcal{B} = \boldsymbol{b}) \sim \mathcal{N}(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{b} + \boldsymbol{o}, \sigma^2 \boldsymbol{W}^{-1})$$

$\mathcal{B} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$ a parameterized $q \times q$ variance-covariance matrix, $\boldsymbol{\Sigma}$,

As a variance-covariance matrix, $\boldsymbol{\Sigma}$ must be positive semidefinite.

$$\boldsymbol{\Sigma}_{\boldsymbol{\theta}} = \sigma^2 \boldsymbol{\Lambda}_{\boldsymbol{\theta}} \boldsymbol{\Lambda}_{\boldsymbol{\theta}}^{\top},$$

a *relative covariance factor*, $\boldsymbol{\Lambda}_{\boldsymbol{\theta}}$, which is a $q \times q$ matrix,

Final Remarks:

- This formulation of linear mixed models allows for a relatively compact expression for the profiled log-likelihood of θ

- The matrices associated with random effects, $\mathbf{Z}$ and $\mathbf{\Lambda}_{\theta}$, typically have a sparse structure with a sparsity pattern that encodes various model assumptions.

- More on the details on the structure and how to represent it efficiently can be found in the assigned readings, R demo, or the live sessions.

- The setup above is very general. The interface provided by *lme4*'s *lmer* function is less general than the model described.

- To take advantage of the entire range of possibilities, one may have to use the modular functions or explore the experimental *flexLambda* branch of *lme4* on *Github*

Berkeley SCHOOL OF INFORMATION