A Novice's Guide to Understanding Mixed Effects Models

Katie Drager
University of Hawaiʻi at Mānoa

Jennifer Hay
University of Canterbury

Mixed Effects Models

- A growing number of sociolinguists are using mixed effects models to analyse their data statistically.

- This talk will step through the basics for those new to mixed effects modeling who:
  - would like to adopt mixed effects techniques, or
  - would simply like to interpret results from studies using these techniques

Mixed Effects Models: what are they anyway?

- A type of regression model (like Varbrul) that takes into consideration variation that is not generalisable to the independent variables (unlike Varbrul)
- Such variation may include variation across different speakers, or different words.

- Allows:
  - continuous dependent and independent variables (e.g., Euclidean distance, speech rate)
  - interactions between any combination of discrete and continuous variables

Choosing your predicting factors

- fixed effects: the independent variables that would normally be included in sociolinguistic analyses
    - e.g., age, gender, phonological environment

- random effects: the variables that are specific to your data sample
    - e.g., speaker, word, listener, item

If your dependent variable is…

- Continuous: use a linear regression model with mixed effects
- Binary: use a logistic regression model with mixed effects

R: not just a sound

- Statistical package that is available free online:

  http://www.r-project.org/

- Uses command line interface, so can be intimidating for new users. But it's up-to-date, gives you total control, and it's not as hard as it looks.

- As both authors use R exclusively, we will be using R for this demonstration.

- For mixed effects models in R, use the lme4 package.

Steps to Fit a Mixed Effects Model

1. Set up data in spreadsheet in a way that R can interpret it.
2. Read data in to R.
3. Optional: subtract mean from continuous variables
4. Plot your data, and perhaps even run other types of analyses (e.g., CART) to get a better picture of the trends in your data.
5. Try out a model. Include potential interactions.
6. Include factors as independent variables in the model if they reach significance or if they are included as a control variable (e.g., phonological environment).

Example of fitting a model

Case study: Drager, Hay & Walker (forthcoming)

- a phonetic accommodation study in New Zealand, investigating the relationship between KIT production, exposure to “good” and “bad” facts about Australia, and whether the speaker is a sports fan.

- KIT is raised in the speech of Australians and centralised in the speech of New Zealanders.

Task

1. read wordlist

2. read fact lists:

   – all read facts about orchids and zebras: 1 group only read these facts (control)

   – 1 group read “good” facts about Australia

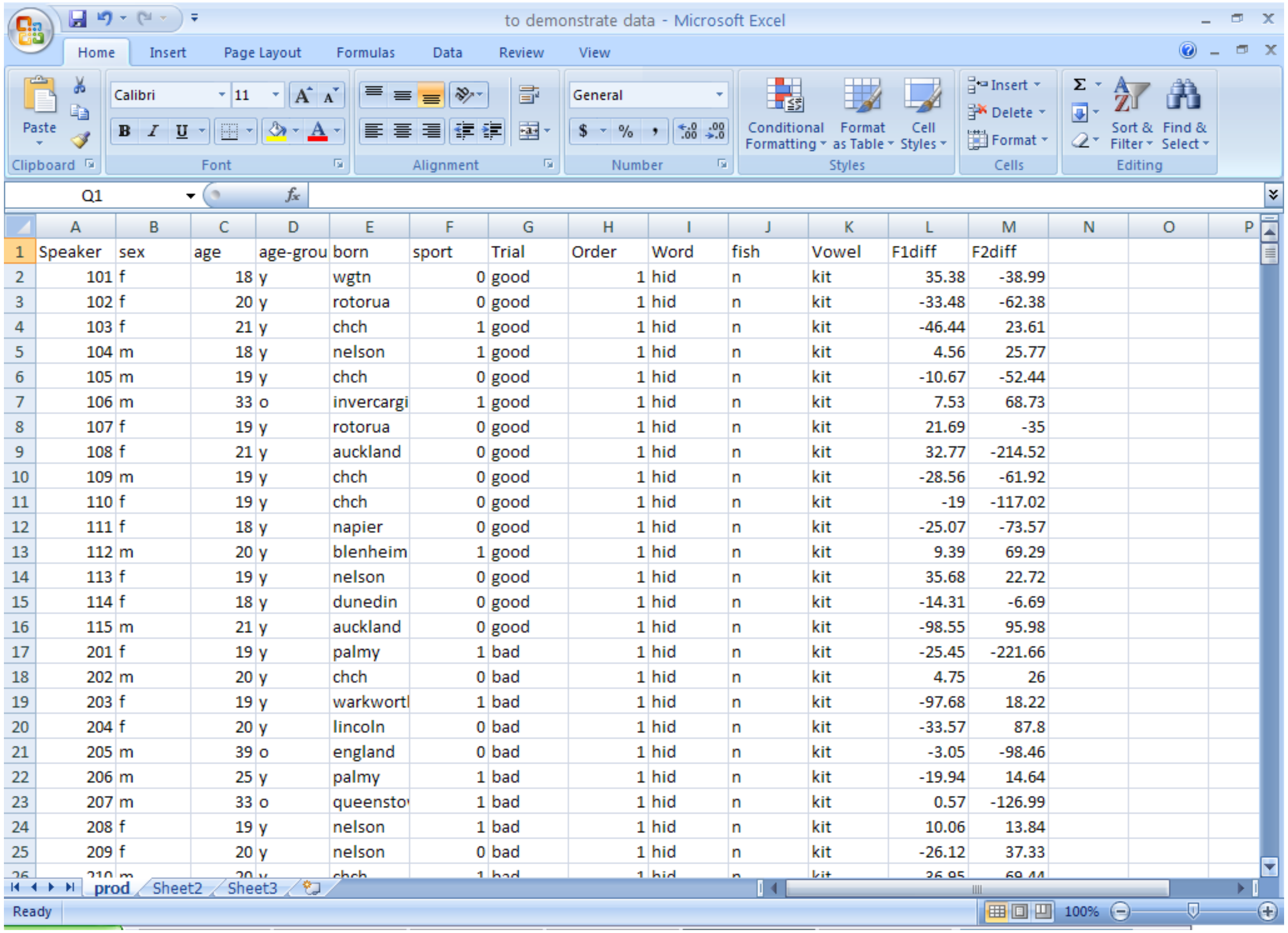
     The Australian government’s donation of $1 billion dollars to the Tsunami relief effort was the biggest made by any country, including those with considerably bigger populations.

   – 1 group read “bad” facts about Australia

     As of 2005, Australia was the world's largest emitter per capita of greenhouse gases, and has still not signed the Kyoto Protocol.

3. read wordlist again

Data



Speaker	sex	age	age-grou	born	sport	Trial	Order	Word	fish	Vowel	F1diff	F2diff
101	f	18	y	wgtn	0	good	1	hid	n	kit	35.38	-38.99
102	f	20	y	rotorua	0	good	1	hid	n	kit	-33.48	-62.38
103	f	21	y	chch	1	good	1	hid	n	kit	-46.44	23.61
104	m	18	y	nelson	1	good	1	hid	n	kit	4.56	25.77
105	m	19	y	chch	0	good	1	hid	n	kit	-10.67	-52.44
106	m	33	o	invercargi	1	good	1	hid	n	kit	7.53	68.73
107	f	19	y	rotorua	0	good	1	hid	n	kit	21.69	-35
108	f	21	y	auckland	0	good	1	hid	n	kit	32.77	-214.52
109	m	19	y	chch	0	good	1	hid	n	kit	-28.56	-61.92
110	f	19	y	chch	0	good	1	hid	n	kit	-19	-117.02
111	f	18	y	napier	0	good	1	hid	n	kit	-25.07	-73.57
112	m	20	y	blenheim	1	good	1	hid	n	kit	9.39	69.29
113	f	19	y	nelson	0	good	1	hid	n	kit	35.68	22.72
114	f	18	y	dunedin	0	good	1	hid	n	kit	-14.31	-6.69
115	m	21	y	auckland	0	good	1	hid	n	kit	-98.55	95.98
201	f	19	y	palmy	1	bad	1	hid	n	kit	-25.45	-221.66
202	m	20	y	chch	0	bad	1	hid	n	kit	4.75	26
203	f	19	y	warkwortl	1	bad	1	hid	n	kit	-97.68	18.22
204	f	20	y	lincoln	0	bad	1	hid	n	kit	-33.57	87.8
205	m	39	o	england	0	bad	1	hid	n	kit	-3.05	-98.46
206	m	25	y	palmy	1	bad	1	hid	n	kit	-19.94	14.64
207	m	33	o	queenstoı	1	bad	1	hid	n	kit	0.57	-126.99
208	f	19	y	nelson	1	bad	1	hid	n	kit	10.06	13.84
209	f	20	y	nelson	0	bad	1	hid	n	kit	-26.12	37.33
210	m	20	y	chch	1	bad	1	hid	n	kit	36.95	69.44

the model

- Dependent variable = F2diff:
    - F2 of the first reading of a token – F2 of the second reading of the token (after being exposed to the facts)
- Independent variables
    - F1 in first token
    - F2 in first token
    - Condition (good facts, bad facts, control)
    - Sports fan vs. not
- Interaction between condition and sports-fandom

DV

IVs

random effects

```
> test = lmer(normF2diff ~ normF1first + normF2first + Trial * Sport.no  + (1|subject)+(1|Word), data=kitprod)
> pvals.fnc(test)
$fixed
                        Estimate MCMCmean HPD95lower HPD95upper  pMCMC Pr(>|t|)
(Intercept)              -0.0762  -0.0794    -0.1966     0.0225 0.1144   0.1007
normF1first               0.0638   0.0765     0.0136     0.1357 0.0162   0.0372
normF2first               0.2772   0.2173     0.0881     0.3505 0.0016   0.0000
Trialcontrol              0.1195   0.1220     0.0358     0.2100 0.0064   0.0356
Trialgood                 0.1373   0.1231     0.0197     0.2239 0.0160   0.0351
Sport.non                 0.0936   0.0830    -0.0025     0.1706 0.0588   0.0873
Trialcontrol:Sport.non   -0.1753  -0.1673    -0.3059    -0.0231 0.0234   0.0564
Trialgood:Sport.non      -0.1823  -0.1653    -0.2961    -0.0370 0.0106   0.0279

$random
    Groups        Name Std.Dev. MCMCmedian MCMCmean HPD95lower HPD95upper
1  subject (Intercept)   0.0849     0.0504   0.0494     0.0151     0.0818
2     Word (Intercept)   0.0554     0.0571   0.0733     0.0000     0.1839
3 Residual               0.1332     0.1431   0.1435     0.1270     0.1605

> 
```

Steps to Interpreting the Output

1. Look at estimated intercept.
   - If DV is continuous: the intercept is the estimated value of the DV if all IVs are the defaults.
   - If DV is binary: the intercept is the log odds of the DV being one factor rather than the other.

The estimated intercept is negative, which indicates a larger F2 value in the second reading (a fronter vowel)

```
> test = lmer|normF2diff ~ normF1first + normF2first + Trial * Sport.no  + (1|subject)+(1|Word), data=kitprod)
> pvals.fnc(test)
$fixed
                        Estimate MCMCmean HPD95lower HPD95upper  pMCMC Pr(>|t|)
(Intercept)              -0.0762  -0.0794    -0.1966     0.0225 0.1144   0.1007
normF1first               0.0630   0.0765     0.0136     0.1357 0.0162   0.0372
normF2first               0.2772   0.2173     0.0881     0.3505 0.0016   0.0000
Trialcontrol              0.1195   0.1220     0.0358     0.2100 0.0064   0.0356
Trialgood                 0.1373   0.1231     0.0197     0.2239 0.0160   0.0351
Sport.non                 0.0936   0.0830    -0.0025     0.1706 0.0588   0.0873
Trialcontrol:Sport.non   -0.1753  -0.1673    -0.3059    -0.0231 0.0234   0.0564
Trialgood:Sport.non      -0.1823  -0.1653    -0.2961    -0.0370 0.0106   0.0279

$random
    Groups        Name Std.Dev. MCMCmedian MCMCmean HPD95lower HPD95upper
1  subject (Intercept)   0.0849     0.0504   0.0494     0.0151     0.0816
2     Word (Intercept)   0.0554     0.0571   0.0733     0.0000     0.1839
3 Residual               0.1332     0.1431   0.1435     0.1270     0.1605

> 
```

Steps to Interpreting the Output

2. Look at IVs’ estimated coefficients.
   – Positive vs. negative
   – How close to zero (= more positive or more negative)
   – To determine estimated value for DV given non-default values for a categorical IV, add IV’s coefficient to the intercept’s estimated coefficient.
   – To determine estimated value for DV for a continuous IV, add value of IV * IV’s coefficient to the intercept’s estimated coefficient.

The estimate for the first token’s F2 is positive. This means that the greater the first token’s F2 was, the smaller the F2 was in the second token (more centralization between readings of the words).

```
> test = lmer(normF2diff ~ normF1first + normF2first + Trial * Sport.no  + (1|subject)+(1|Word), data=kitprod)
> pvals.fnc(test)
$fixed
                       Estimate MCMCmean HPD95lower HPD95upper  pMCMC Pr(>|t|)
(Intercept)             -0.0762  -0.0794    -0.1966     0.0225 0.1144   0.1007
normF1first              0.0638   0.0765     0.0136     0.1357 0.0162   0.0372
normF2first              0.2772   0.2173     0.0881     0.3505 0.0016   0.0000
Trialcontrol             0.1195   0.1220     0.0358     0.2100 0.0064   0.0356
Trialgood                0.1373   0.1231     0.0197     0.2239 0.0160   0.0351
Sport.non                0.0936   0.0830    -0.0025     0.1706 0.0588   0.0873
Trialcontrol:Sport.non  -0.1753  -0.1673    -0.3059    -0.0231 0.0234   0.0564
Trialgood:Sport.non     -0.1823  -0.1653    -0.2961    -0.0370 0.0106   0.0279

$random
    Groups        Name Std.Dev. MCMCmedian MCMCmean HPD95lower HPD95upper
1  subject (Intercept)   0.0849     0.0504   0.0494     0.0151     0.0818
2     Word (Intercept)   0.0554     0.0571   0.0733     0.0000     0.1839
3 Residual               0.1332     0.1431   0.1435     0.1270     0.1605

> |
```

This is significant (p<.0001), as indicated by the p-value.

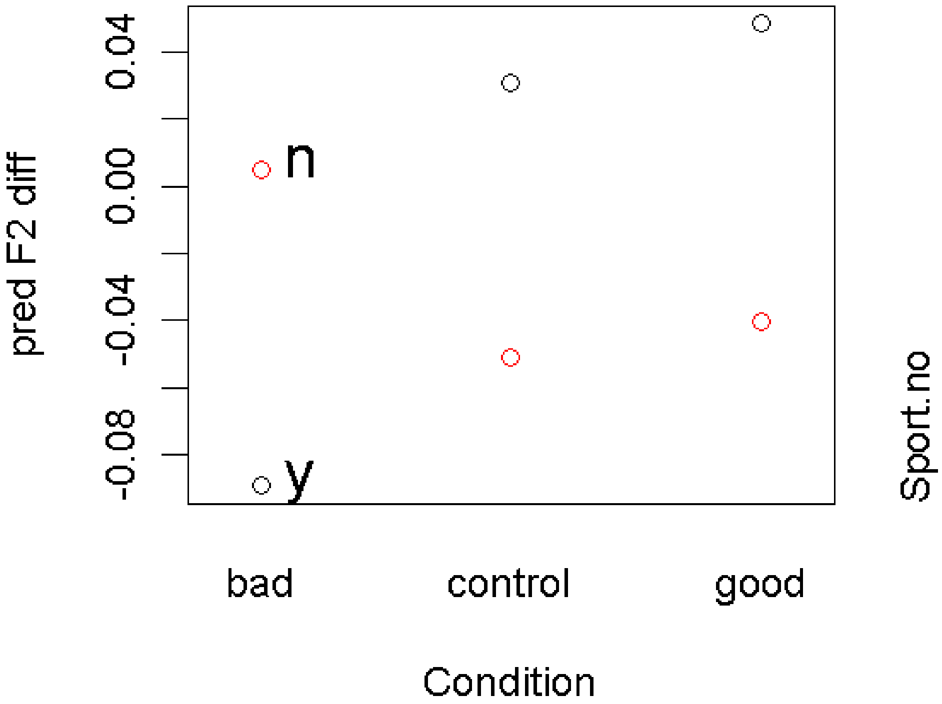
There is an interaction between the condition (good facts or bad facts) and whether the speaker was a sports fan. This needs to be understood within the context of the estimated coefficients when they are not interacting and when they are.

```
> test = lmer(normF2diff ~ normF1first + normF2first + Trial * Sport.no  + (1|subject)+(1|Word), data=kitprod)
> pvals.fnc(test)
$fixed
                        Estimate MCMCmean HPD95lower HPD95upper   pMCMC Pr(>|t|)
(Intercept)              -0.0762  -0.0794    -0.1966     0.0225 0.1144   0.1007
normF1first               0.0638   0.0765     0.0136     0.1357 0.0162   0.0372
normF2first               0.2772   0.2173     0.0881     0.3505 0.0016   0.0000
Trialcontrol              0.1195   0.1220     0.0358     0.2100 0.0064   0.0356
Trialgood                 0.1373   0.1231     0.0197     0.2239 0.0160   0.0351
Sport.non                 0.0936   0.0830    -0.0025     0.1706 0.0588   0.0873
Trialcontrol:Sport.non   -0.1763  -0.1673    -0.3059    -0.0231 0.0234   0.0564
Trialgood:Sport.non      -0.1823  -0.1653    -0.2961    -0.0370 0.0106   0.0279

$random
    Groups        Name Std.Dev. MCMCmedian MCMCmean HPD95lower HPD95upper
1  subject (Intercept)   0.0849     0.0504   0.0494     0.0151     0.0818
2     Word (Intercept)   0.0554     0.0571   0.0733     0.0000     0.1839
3 Residual               0.1332     0.1431   0.1435     0.1270     0.1605

> |
```

The interaction between facts and sports-fandom: fans and non-fans responded differently depending on whether they read good or bad facts, an interaction which is significant (p<.05).

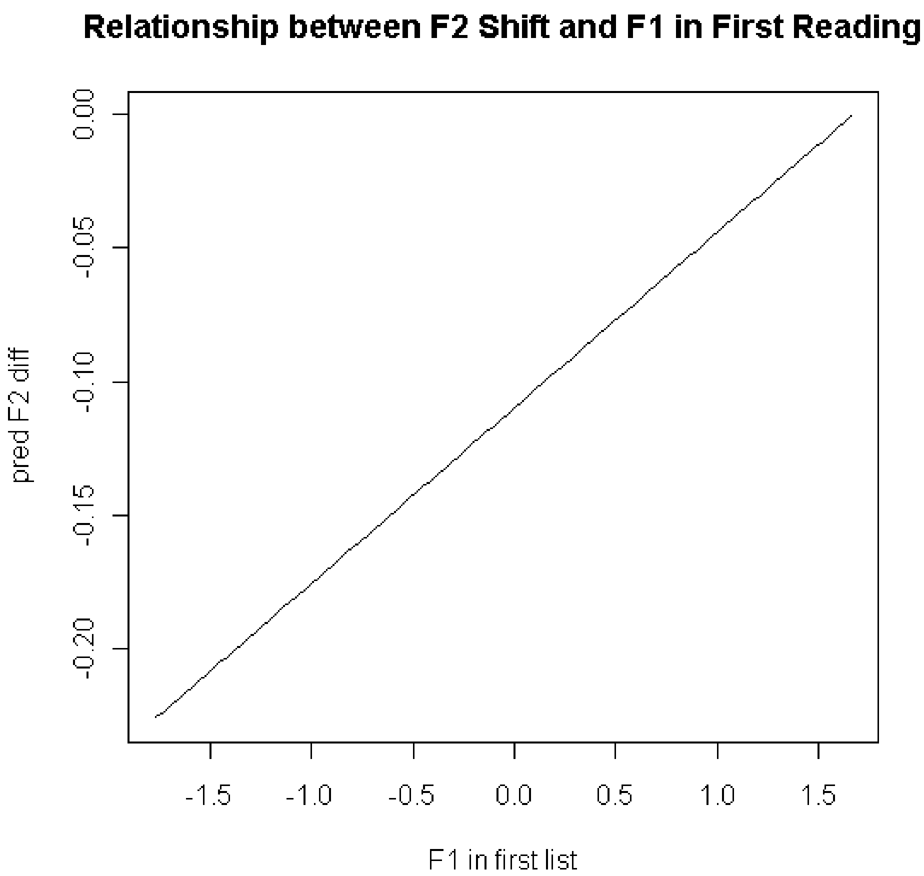
Plotting Graphs

```
par(cex=1.5)
plotLMER.fnc(test, pred="Trial", intr=list("Sport.no", c("y", "n"), "beg", list(c("black", "red"), rep(1,2))), xlabel = "Condition", ylabel= "pred F2 diff", cex=1.5)
```



Plotting Graphs

```
plotLMER.fnc(test, pred = "normF1first", xlabel = "F1 in first list", ylabel = "pred F2 diff")
title("Relationship between F2 Shift and F1 in First Reading")
```



Things to look out for

1. Extremely large or small coefficients are indicative of a model that is overfit (i.e., there are too many independent variables included in the model so that the variation is overspecified).

2. Correlated variables.

3. False negatives.

4. For factors with more than 2 levels, be sure to set the 'default' level appropriately so you can assess the significance of the most relevant contrasts.

5. Numbers will be interpreted by R as continuous, even for factors where that makes no sense (e.g., subjects labeled as 1, 2, 3, 4…). To make sure that R does not treat these factors as continuous, use:

   ```
   ausprod$subject = as.factor(ausprod$Participant.number)
   ```

Interpreting Random Effect Intercepts

- The intercepts for random effects can be interpreted in much the same way as the fixed effects' estimated coefficients
  - Are they positive or negative?
  - How close to zero?
  - Add to the model's estimated intercept to see estimated value of DV for that speaker or word.


```
> ranef(test)
$subject
     (Intercept)
101  0.067754636
102 -0.090714863
103 -0.043169863
104  0.038772122
105  0.075030274
106 -0.072001454
107  0.087607373
108 -0.109799756
109  0.053378468
110  0.033699791
111 -0.072466839
112  0.076399195
113  0.018033886
114  0.015237093
115 -0.077760064
201 -0.081429032
307 -0.071457691
308 -0.017039804
309  0.131317008
310  0.024073909

$Word
     (Intercept)
fish  0.05831931
hid  -0.05522102
hint  0.02587080
hit  -0.02896910
```

random effect intercepts for each subject

random effect coefficients for each word

Interpreting Random Effect Intercepts

- Positive values = an increase in the DV for that speaker or word, given the IVs included in the model.

- Negative values = a decrease in the DV for that speaker or word, given the IVs included in the model.

- Closer to zero = deviate less from the model's intercept


```
> ranef(test)
$subject
     (Intercept)
101   0.067754636
102  -0.090714863
103  -0.043169863
104   0.038772122
105   0.075030274
106  -0.072001454
107   0.087607373
108  -0.109799756
109   0.053378468
110   0.033699791
111  -0.072466839
112   0.076399195
113   0.018033886
114   0.015237093
115  -0.077760064
201  -0.081429032
307  -0.071457691
308  -0.017039804
309   0.131317008
310   0.024073909

$Word
      (Intercept)
fish   0.05831931
hid   -0.05522102
hint   0.02587080
hit   -0.02896910
```

random effect intercepts for each subject

random effect coefficients for each word

Why Use Mixed Effects Models?

Statistically rigorous

Increasingly adopted by linguists in other subfields

Allows for continuous data and easy testing of interactions

Can exploit the random intercept values for various purposes in sociolinguistic analyses

(Drager & Hay, under review)

Mixed effects models…

…one more (very sharp) tool in our sociolinguistic toolbox

Acknowledgments

- Abby Walker
- Sali Tagliamonte for organising this workshop