Mixed effects regression II

What we've seen so far (1/)

- Some independent variables are *random effects*, categorical variables sampled from a large or even infinite set of possible values (e.g., subject; "language"/stimulus/item/word/sentence).
- When these are *nested* with other covariates (e.g., subject and subject age, education, social class; stimulus and fixed properties of that stimulus like length and frequency), this may introduce multicollinearity.

What we've seen so far (2/)

- Treating random effects as *fixed effects* overfits and requires a huge number of degrees of freedom, yet we are usually uninterested in using them for null hypothesis testing.
- Aggregating observations across random effects underestimates variance and the number of free degrees of freedom; and, it is only straightforward to aggregate across one random effect at a time.
- Omitting random effects altogether gives rise to *omitted variable bias* (e.g., Simpson's paradox).

Mixed effects models

- Provide a parsimonious (albeit internally quite complex) way to estimate linear models which contain random effects.
- *Random intercepts* are estimates like multinomial categorical fixed effects, but they are *shrunk* and are roughly normally distributed with mean $\bar{X} = 0$.

```
RT ~ (1 | Subj) + (1 | Item)
```

- It is straightforward to compare two mixed effects models so long as
  - they have the same random effects structure, and
  - one's fixed effects "nests" (is a proper superset) of the other's.

Outline for today

- *Random slopes*, interactions between fixed and random effects, and
- *generalized* linear mixed effects models such as *logistic mixed effects regression*.

Random slopes

Lexical decision example

The dependent variable is reaction time; independent variables include:

- Fixed effects:
  - `NativeLanguage` ("English" or "other")
  - (word) `Frequency`, and
  - `Trial` (number).
- Random effects:
  - `Subject`
  - `Word`
- Nesting relationships:
  - `NativeLanguage` and `Subject`
  - `Frequency` and `Word`

Lexical decision example (1/)

```
> r1 <- lmer(RT ~ NativeLanguage + Frequency + Trial +
+               (1 | Subject) + (1 | Word), data = d)
```

Lexical decision example (2/)

```
> summary(r1)
...
Fixed effects:
                 Estimate Std. Error t value
(Intercept)      6.606391   0.041596 158.824
NativeLanguage1 -0.076053   0.028839  -2.637
Frequency       -0.043783   0.006140  -7.130
Trial           -0.009123   0.004310  -2.117
...
```

Lexical decision example (3/)

```
> lrtest(update(r1, . ~ . - Trial), r1)
...
  #Df LogLik Df  Chisq Pr(>Chisq)
1   6 472.95
2   7 470.66  1 4.5795    0.03236 *
...
```

Interpretation

Our prediction of a trial's RT is given by summing

- the intercept,
- the coefficient for `NativeLanguage`,
- the coefficient for word `Frequency` times the frequency
- the coefficient for `Trial` number times the trial number,
- the effect for that `Subject`, and
- the effect for that `Word`.

Subjects are different...

The `Subject` random intercept recognizes that some subjects have faster or slower mean RTs than others. But there are other ways subjects can differ:

- e.g., perhaps some subjects get fatigued faster or slower than average.

If `Subject` and `Trial` were both fixed effects, we would simply add a standard interaction term; an interaction between a random effect and a fixed effect is known as a *random slope*.

Lexical decision example (4/)

```
> r2 <- lmer(RT ~ NativeLanguage + Frequency + Trial +
              (1 + Trial | Subject) + (1 | Word), data = d)
```

The portion `(1 + Trial | Subject)` specifies a per-`Subject` random intercept and a per-`Subject` random slope of `Trial`.

So we are estimating the overall correlation between trial number and RT (the fixed effect), but each subject adjusts the slope of that correlation; the RT-given-trial number effect/slope/coefficient for a given subject is given by $\beta_{\text{Trial}} + \beta_{\text{Trial(subj=}n\text{)}}$.

Lexical decision example (5/)

```
> summary(r2)
...
Fixed effects:
                 Estimate Std. Error t value
(Intercept)       6.606391   0.041596 158.824
NativeLanguage1 -0.076053   0.028839  -2.637
Frequency       -0.043783   0.006140  -7.130
Trial           -0.009123   0.004310  -2.117
...
```

Not much change in estimate or standard error, here.

Lexical decision example (6/)

```
> lrtest(update(r2, . ~ . - Trial), r2)
...
   #Df LogLik Df  Chisq Pr(>Chisq)
1    8 486.95
2    9 483.73  1 6.4321    0.01121 *
...
```

However, the χ^2 statistic is larger, and the associated *p*-value is smaller because we are now accounting for within-subject variation in fatigue.

Categorical/categorical random slopes

Usually, the fixed portion of a random slope (an interaction between a fixed and random effect) is continuously-valued; it is difficult to estimate a random slope when the fixed portion is an multinomial categorical effect.

Generalized mixed effects models

Generalizing mixed effects linear regression

With (non-mixed) linear regression we can back off from the assumption that the dependent variable is continuous using *generalized* linear models, which transform the raw prediction using a *link function*.

E.g., *logistic regression* uses the logit function to map from log-odds to probabilities.

This trick can also be comfortably used with mixed effects models; the only thing we need to change is that we need to:

- use the `glmer` function rather than `lmer`, and
- specify a link function (e.g., for logistic regression, `family = binomial`).

Disfluency example (1/)

MacFarlane et al. (2017) use *mixed effects logistic regression* to model disfluency use in a sample of children with or without development disorders.

They hypothesized that the relative frequency of *content mazes* (revisions, repetitions, and false starts) and *filler mazes* (like *uh*, *um*, *I mean*, etc.) would be correlated with developmental disorders.

Repetition—REP	(y){y}z	*(My) {My} dog is nice.*
	mid-word interruption	*My (do-) {dog} is nice.*
Revision—REV	(xy){xz}	*(My dog) {My cat} is nice.*
		She (goes) {went}.
	(xz){yz}	*(He's very) {She's very} friendly.*
[Deletion—DEL]	(xyz){xz}	*(My old dog) {My dog} is nice.*
	(xy){x}	*My dog (likes to) {likes} food.*
[Insertion—INS]	(xz){xyz}	*(My dog) {My yellow dog} is nice.*
	(y){xy}	*My dog likes (the) {all the} food.*
False start—FS	(xyz) abc	*(But what if) My friend likes dogs.*
Filler—F	discourse markers and filled pauses	*like*, *um*, *uh*, *mm*, *hmm*, *I mean*, *ah*

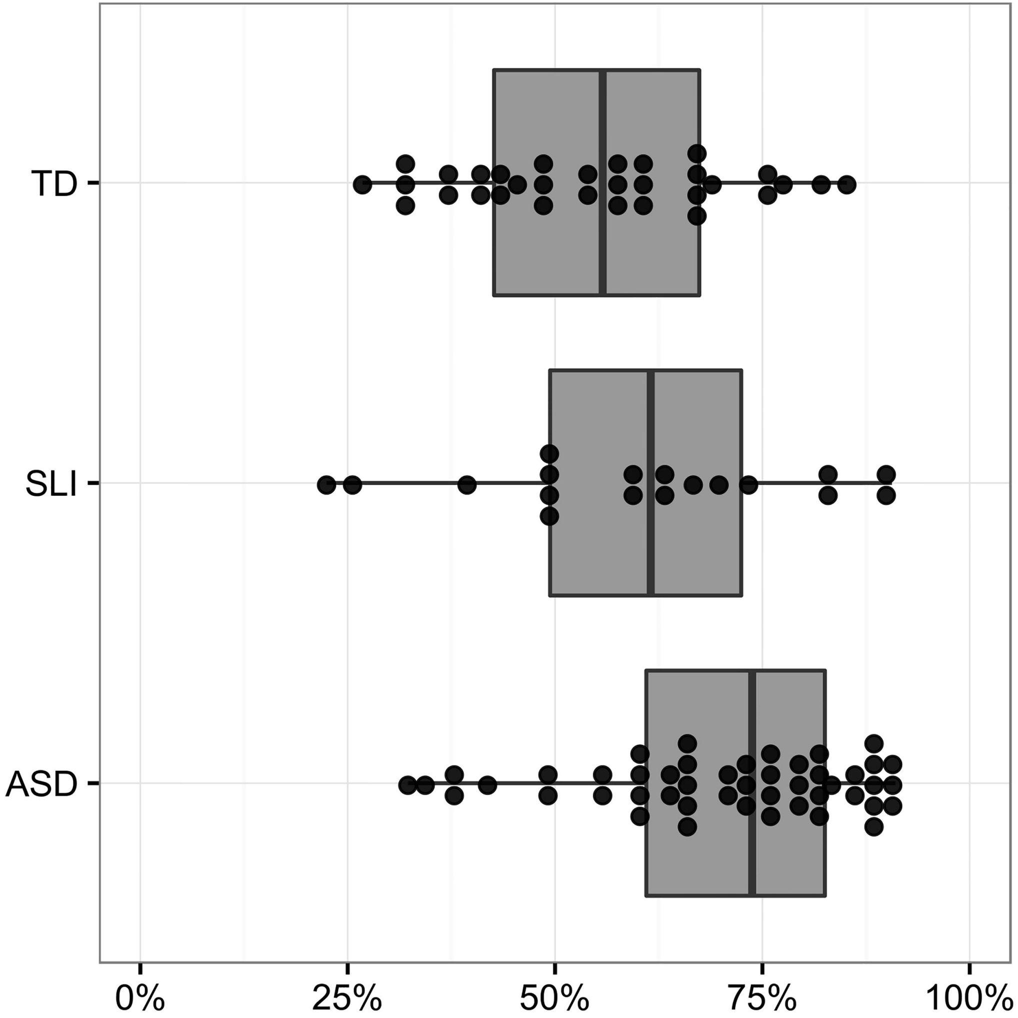
https://doi.org/10.1371/journal.pone.0173936.t001

Disfluency example (2/)

Their dependent variable was binomial: *content* vs. *filler* mazes.

Independent variables included fixed effects of:

- diagnosis (TD: typically developing; ASD: autism spectrum disorder; SLI: specific language impairment),
- verbal IQ, measured using the age-appropriate Wechsler scales,
- and the interview activity (Play, Description of a Picture, Telling Story From A Book, Conversation).


TD
SLI
ASD
0%
25%
50%
75%
100%

	Log-odds	S.E.s	χ^2	$P(\chi^2)$
(Intercept)	0.670	0.09		
Group:			12.18	.002
ASD	0.386	0.11		
SLI	−0.055	0.16		
TD	−0.331	0.15		
VIQ	−0.001	0.09	0.00	.988
ADOS Activity:			120.34	<.001
Play	0.074	0.04		
Description Of A Picture	−0.088	0.04		
Telling A Story From A Book	0.360	0.05		
Conversation	−0.347	0.03		

Mixed effects logistic regression on the relative frequencies of content mazes versus fillers; predictors which favor content mazes have positive log-odds and predictors which favor fillers have negative log-odds. VIQ = verbal IQ.

https://doi.org/10.1371/journal.pone.0173936.t004

Disfluency example (3/)

```
r <- glmer(ContentOrFiller ~ DX + VIQ + Activity +
+            (1 | Subject), data = d, family = binomial)
```

Final notes

Model comparison

Because we cannot use the likelihood ratio test to compare models with different random effects structure, we cannot easily ask "do we need a random intercept/slope?" with null hypothesis testing.

Inestimable models (1/)

It is easy to build a mixed effects model with random effects that cannot be estimated from the data. If so you may see a warning like the following:

```
Warning messages:
1: In checkConv(attr(opt, "derivs"), opt$par, ctrl =
control$checkConv,  :
  Model failed to converge with max|grad| = 0.24984 (tol =
0.002, component 1)
2: In checkConv(attr(opt, "derivs"), opt$par, ctrl =
control$checkConv,  :
  Model is nearly unidentifiable: very large eigenvalue
 - Rescale variables?
```

Inestimable models (2/)

In some cases, this can be solved by scaling (i.e., centering or, preferably, standardizing) all continuous fixed effects and the continuous portion of a random slope (in fact that was necessary to make the `Trial` given `Subject` random intercept converge in the example above).

But if that doesn't work, we have no choice but to omit the random slope, unless we turn to much more complex models.

Recommended reading on mixed effects models

Repeating myself from last week, but:

- Baayen (§7) does a good job both motivating and exemplifying mixed models.
- Johnson (§4.4, §7.3-7.4) and Gorman & Johnson (2014: 226-229) also give a number of worked examples.

(This is all indicated on the course website.)

Questions? Please take them to email, or Slack.