Module 4A: Maximal Random Effects

- According to recent work (see Barr et al., JML, 2013), the best (=most conservative) means to estimate effects is to use the *maximal* random effects structure.

- This means
  - Include all random effects in the model. For example, in an experimental study, not just participants but also items (e.g., if there are multiple words in each experimental condition).
  - For each random effect, include a random intercept and random slopes for every fixed effect that varies with respect to the random effect (and have these be correlated).

Module 4A: Maximal Random Effects

- Example: Lexdec
  - Focus on the model with frequency, native language background, and their interaction
- Two random factors: subjects and words.

- Begin with a model that has all of the fixed effects you're interested in plus random intercepts [don't enter this code into R—this is just to illustrate the procedure]

RT ~ Frequency*NativeLanguage +

(1|Subject)+(1|Word)

Module 4A: Maximal Random Effects

- Inside of your text editor [NOT R], copy and paste <u>the entire fixed effects structure</u> so as to give correlated random slopes for every fixed effect, over every random effect

RT ~ Frequency*NativeLanguage +

(1 + Frequency*NativeLanguage |Subject)+

(1 + Frequency*NativeLanguage |Word)

- Now, ask yourself: does this make sense?
  - For each of the potential random slopes, does this fixed effect vary with the random effect?

Module 4A: Maximal Random Effects

- Start with subjects

RT ~ Frequency*NativeLanguage +

(1 + Frequency*NativeLanguage |Subject)+

(1 + Frequency*NativeLanguage |Word)

- Does frequency vary by subject?
  - Yes; each subject says both high and low frequency words
- Does native language vary by subject?
  - NO. Each subject is either native or non-native.
  - Given that the main effect can't have a random slope, neither can the interaction. This yields:

(1 + Frequency |Subject)

Module 4A: Maximal Random Effects

- Now consider words

RT ~ Frequency*NativeLanguage +

(1 + Frequency|Subject)+

(1 + Frequency*NativeLanguage |Word)

- Does frequency vary by word?
  - NO. Each word has a given frequency.
  - Given that the main effect can't have a random slope, neither can the interaction.
- Does native language vary by word?
  - Yes. Both native and non-native speakers respond to each word.

This yields:

(1 + NativeLanguage|Word)

Module 4A: Maximal Random Effects

- To apply this, first execute code chunk 28. This re-codes Native Language in the lexdec data set using contrasts, and then translates that contrast into a new numeric column in the lexdec dataframe.

Module 4A: Maximal Random Effects

- Maximal random effect structure given in code chunk 29
  - Subjects: Random intercept, correlated slope for frequency (each person sees high and low frequency words).
    - No random slope for native language background (subjects are either native or not)
  - Words: Random intercept, correlated slope for native language background (each word is said by each participant group)
    - No random slope for frequency (words have some frequency—they don't vary with respect to this).

Module 4A: Maximal Random Effects

- Code chunk 30 tests whether the interaction is significant, using our subtraction method

- To communicate finding:
  – There was a significant interaction of frequency and native language background, such that non-native speakers' reaction times showed a stronger influence of word frequency (β = -0.027, SE β = 0.01, $\chi^2(1) = 6.94$, $p < .05$)

Module 4A: Maximal Random Effects

- **Question 15. Use the subtraction method to test for the significance of the main effects of frequency and native language background. Report the results of your tests.**

Module 4B: Application to Continuous Data

- For this module, you'll need to use the dataset in the file neighbors.txt. This provides voice onset times (VOTs) for an experiment reported in Baese-Berk & Goldrick (2009; this paper used a different set of statistical techniques).

- In this experiment, participants produced words that begin with the voiceless stops /p, t, k/. Some of these words form a *minimal pair* with another English word that differs only in voicing. For example, *cod* forms a minimal pair with the word *god*. Each of these words is matched (or paired) with a similar word that doesn't form a minimal pair. For example, *cod* is paired with *cop*, which has no minimal pair *gop*.

Module 4B: Application to Continuous Data

- These words were produced in a direction-giving task. A speaker and a hearer looked at two different monitors with the same three words (e.g., *cod lamp yell*). A speaker told the hearer which of three words the hearer should click on (e.g., *Click on the cod*). Words that form minimal pairs appeared with (e.g., *cod god yell*) or without (e.g., *cod lamp yell*) their minimal pair counterpart.
- This yields three conditions:
  - No Competitor (e.g., *cop*);
  - Competitor-Absent (e.g., *cod lamp*);
  - Competitor-Present (e.g., *cod god*)

Module 4B: Application to Continuous Data

- The data fields in the file are:
- Subject Subject ID
- Condition (as on previous slide)
- VOT Average voice onset time across three trials
- Pair Pair ID (e.g., *cop* and *cod* have the same Pair ID)
- Consonant Initial consonant of the word (p/t/k)

Module 4B: Application to Continuous Data

- **Question 16a. Build a linear mixed regression model with the maximal random effects structure predicting VOT using contrast-coded factors representing Condition and Consonant (leave off interactions between these factors).**
  - **Code Condition using two contrasts: (i) Competitor conditions, as a group, vs. the No Competitor condition; (ii) Competitor-Present vs. Competitor-Absent.**
    - Note: code competitor as negative, no competitor as positive; competitor present as negative, absent as positive (if not you will get an error—more later).
  - **Code Consonant using two contrasts: (i) lingual, as a group (/t/ and /k/), vs. bilabial (/p/); (ii) alveolar vs. velar. Note this is coded differently than in the votPOA dataset examined earlier.**

Module 4B: Application to Continuous Data

- **Question 16b. Test for the significance of the fixed effects factors included in your model.**

- **Question 16c. Examine the residuals of the model. If they are non-normal, how might you address this?**

Module 4C: Application to Categorical Data

- When your experiment gets complex, your model gets complex as well.
- Consider Bradlow and Alexander (2007; this paper used a different set of statistical techniques). Prior to this work, it had been shown that native speakers' ability to accurately perceive speech in noise benefits both from clear speech style (i.e., the "exaggerated" speech produced in difficult listening environments) as well as semantic context (e.g., when words are presented in a predictable sentence as in *A stitch in time saves <u>nine.</u>*). Previous work with non-native listeners had shown that they have more difficulty understanding speech in noise than native listeners. Furthermore, such listeners do benefit from clear speech. However, when speech was presented in a plain (i.e., unexaggerated) speech style, non-native listeners showed no benefit from semantic context.

Module 4C: Application to Categorical Data

- The purpose of Bradlow and Alexander (2007) was to examine if semantic context would enhance non-native speech perception when combined with the good acoustic cues present in clear speech. Sentences were read by a native English speaker in both plain (conversational) and clear speech styles. Half of the sentences strongly predicted the final word (e.g., "The color of a lemon is *yellow*."); the other half did not strongly predict the final word(e.g., "Mom thinks that it is *yellow*."). These were presented to native and non-native listeners in noise. The listeners' task was to identify the final word of the sentence.

Module 4C: Application to Categorical Data

- The data are in the file scnn.txt, available on blackboard; it contains 5 columns:
- <u>Subject</u> Subject ID
- <u>Background</u> Language background: Native or Non-Native
- <u>Probability</u> Predictability of final word: Low or High
- <u>Style</u> Style of speech of sentence: Plain or Clear
- <u>Correct</u> Accurate (1) or inaccurate (0)

Module 4C: Application to Categorical Data

- The critical question for these data is: do non-native listeners—unlike natives—show an effect of semantic predictability in clear speech but not in plain speech?

- This is captured by a three-way interaction
  - Non-natives show a two-way interaction between predictability and speech style (with no effect in plain speech)
  - Natives do not show the same two-way interaction (effects in both plain and clear).
  - We're looking for a difference in the size of the two-way interaction across groups; hence, a 3 way interaction.

Module 4C: Application to Categorical Data

- The maximal random effects structure for the Bradlow and Alexander data has a huge number of parameters.
- For each person, you should fit
  - A random intercept
  - A random slope for the 3 main effects
  - A random slope for the 3 two-way interactions
  - A random slope for the three-way interaction
  - Correlations between each pair of the above

Module 4C: Application to Categorical Data

- The lme4 algorithm sometimes just breaks down when dealing with this complexity.

- In this case, if you fit the maximal model, you'll get an error

```
 Warning message:
 In mer_finalize(ans) : iteration limit
reached without convergence (9)
```

- If you ever see this DO NOT TRUST THE RESULTING LMER OBJECT.

Module 4C: Application to Categorical Data

- What this means is: the algorithm can't fit the parameters.
  - Whatever it gave you as an answer is not likely to be correct.

- How can you help the algorithm out?
  - If you have very small effect sizes, this can happen because the model coefficients are close to R's limits of precision.
  - In this case, it sometimes helps to divide your predictor (numerically coded, of course) by 10 or 100 or 1000. This inflates the size of your coefficients (if the coefficient gets tinier, you have to multiply by it by a larger number).
  - Frequently this is of no help.

Module 4C: Application to Categorical Data

- What you typically must do is reduce the complexity of the model.
  - This gets into the incredibly thorny and complex issue of *model selection*—figuring out what your model structure should be post-hoc (after the data has been gathered).
  - Two basic approaches
    - 'Forward': Start simple and then try more complex models, stopping when lme4 breaks.
    - 'Backward': Start complex and then move to more simple models, stopping when lme4 starts working.
- I favor backward model selection
  - Broadly speaking, including more random slopes is more conservative (=less likely to yield significant results when there is actually no effect).

Module 4C: Application to Categorical Data

- My recipe. Start at 1; if it doesn't converge, go to next step; if converges, stop.
  1. Fit the maximal random effects structure
  2. Drop all correlation parameters (simulations suggest dropping these doesn't make analysis drastically less conservative
  3. Eliminate random slopes for interactions, ordered by complexity (e.g., 3-way, than 2-way)
  4. Eliminate all random slopes.
  5. Eliminate by-item random intercepts
  6. If still won't converge…you will just need to do a multiple regression.

Challenges to Complex Designs

- For the Bradlow and Alexander data set, this procedure stops after all random slopes for interactions have been eliminated.
- Code chunk 30 illustrates this procedure's application. Don't execute those big models that don't converge—unless you have some free time to sit and watch R spin.

Challenges to Complex Designs

- **Question 17. Using the model that DOES converge (last model in chunk 30), perform model comparison to determine the significance of each predictor.**

INTERIM SUMMARY

- Maximal random effects structure: Attempt to make fixed effects vary by each random effects factor, including their correlations.

- When that fails, simplify your random effects structure until model converges.