

Module 3A: Mixed Models Basics

- For the next few slides, we will review a chapter introducing mixed models from the following text:
 - Shravan Vasishth and Michael Broe. (2010) *The Foundations of Statistics: A Simulation-based Approach*.
 - Chapter 7: Available at:
 - http://link.springer.com/content/pdf/10.1007%2F978-3-642-16313-5_7.pdf
- Read Vasishth & Broe excerpt to top of p.152.
 - Code in this chapter is provided at
 - <http://www.ling.uni-potsdam.de/~vasishth/Misc/VasishthBroebook.R>
 - NOTE. To draw some of their awesome graphs, you need to have installed the lattice package. If you did not do that, you can't run chunk 247, 248; that's ok, it just produces the graphs you see in the book.

Module 3A: Mixed Models Basics

- Re-run code on p. 151. (execute chunk 239 before executing 249, 250; this loads in the dataset mathachieve.txt—available from <http://www.ling.uni-potsdam.de/~vasishth/book.html>).
- **Question 12. Calculate the difference between each individual school's slope for SES and the slope of the simple linear model ($lm0$, defined on p. 146 [see chunk 240]). Run a t-test to test the null hypothesis that these differences are equal to 0.**
 - Note: to subtract some number a from every element of a vector x , use the code $x - a$.
- **How does this t-test show that the simple linear regression of achievement scores on SES fails to characterize ?**

Module 3A: Mixed Models Basics

- Discuss up through the top of p. 156.
- **Question 13.** Recall the lexical decision dataset lexdec.
- **13a.** Explain what the terms in equation 7.1 (p. 146) correspond to in a regression model predicting RT from frequency.

Module 3A: Mixed Models Basics

- **13b. Consider a new regression model that allows individual subjects to have different overall differences in reaction times and be differentially affected by frequency. This model would be defined by an equation like 7.4 (p. 152) [although note that frequency isn't centered].**
- **Explain what the terms in 7.4 would correspond to in such a model.**

Module 3A: Mixed Models Basics

- **13c. Let us augment the model in (2b) to allow native English speakers and non-native speakers ('other' language speakers) to have overall RT differences and allow the effect of frequency to vary across native and non-native speakers. This would correspond to equation 7.13 (p. 155).**
- **Explain what the terms in 7.13 would correspond to in such a model.**

Module 3A: Mixed Models Basics

- Read up through 158.
 - NOTE. The lmer code on p. 156 does not match the model description preceding it. It's missing an interaction of SES and Sector. It should be
 - `lme1.fm <- lmer(MathAch ~ SES * Sector + (1 + SES | School), MathScores)`
- Note: we'll explain the statistical test performed by the `anova()` command on p. 158 in just a bit.
- **Question 14a. Use lme4 to construct a model of the lexical decision data including an interaction between Native Language and Frequency.**
- **Question 14b. Is the fit of this model better than one in which the effect of frequency does not vary across subjects? Note: this will be the single difference between the models. Provide quantitative measures to support your claim.**

Module 3A: Mixed Models Basics

- Just as with glm, you can perform logistic regression by adding the argument family="binomial" to the lmer command.
- The output will report Wald z statistics (just as with glm).

Module 3A: Mixed Models Basics

- In lme output (and code—more later), there is a distinction between *fixed* and *random* effects.
- Our simple linear models have a random, normally distributed error term
 - $\varepsilon_i \sim N(0, \sigma^2)$
- For a random *effect*, we assume that the different levels of an *effect* are also normally distributed
 - With a mean of 0 and a variance estimated from the data.
- Makes sense when our effect is a random sample from a much larger population
 - Population of potential subjects
 - Set of possible words

Module 3A: Mixed Models Basics

- Fixed effects are not assumed to be drawn from a distribution.
 - Each level is fit independently and is not constrained to be in some relationship with other levels.
 - More precisely, in dummy coding, each level is assigned a separate coefficient—and each coefficient is not constrained to be in a particular relationship with other coefficients.
- Makes sense when our levels are not random samples
 - They completely define the space of possibilities (e.g., native English vs. not native English)
 - They are not a random sample, so we can't generalize to unseen observations (e.g., learning about labial consonants won't necessarily tell us how alveolars and velars should behave—they are produced with different articulators).

Module 3A: Mixed Models Basics

- In the lme4 command, random effects are specified using the pipe operator |.
- (1|factor-X) specifies a random intercept for different levels of factor X
- (0+factor-A|factor-X) specifies a random slope for factor A for different levels of factor X which is uncorrelated from the random intercept for factor X.
- (1+factor-A|factor-X) specifies a random slope for factor A for different levels of factor X which is correlated with the random intercept for factor X.

Module 3A: Mixed Models Basics

- Note! Just as you have multiple fixed effects, you can have multiple random effects! (like subjects AND items)
- **This is a BIG deal.** This is not something that other multiple regressions or ANOVAs can do.
- These random effects can be “crossed” as well as “nested”
 - Nested: each subject sees all items in all conditions
 - Crossed: subjects see all items, but not in all conditions

Module 3A: Mixed Models Basics

- Example output from book, p. 157:

```
Random effects:
  Groups   Name                Variance Std.Dev. Corr
  School   (Intercept)            3.96385  1.99094
           SES                0.43431  0.65902  0.550
  Residual                        36.80088  6.06637
Number of obs: 7185, groups: School, 160
```

- The intercepts for schools are drawn from a population with mean 0 and an estimated variance of 3.96.
- The school-specific SES slopes are drawn from a population with mean 0 and an estimated variance of 0.43.
- These are correlated at 0.55

Module 3A: Mixed Models Basics

- In the lme command, the fixed effects are all of the factors that occur outside of the (|X) part of the model statement.
- E.g., Bolded factors are fixed effects
- `lmer(MathAch~SES+Sector+(1+SES|School),MathScores)`

Module 3B: Significance Testing

- For continuous variables like RT, the output of lme4 provides effects estimates, standard errors and t values but *no p values*. Why?
 - It's not clear how to estimate the degrees of freedom.
- Remember that the t distribution approaches the z distribution as degrees of freedom increases.
 - For t-tests for simple regressions, the degrees of freedom is on the order of the number of observations.
 - Good rule of thumb: If you have more than 50 observations, $t \approx z$; $|t| > 1.96$ means two-tailed $p < .05$.
 - Some researchers make this assumption, and report p-values assuming $t \approx z$.

Module 3B: Significance Testing

- However:
 - Basis for this assumption is not entirely clear.
 - For logistic regressions, Wald z statistics are known to be biased, particularly for small sample sizes
- Better: Utilize model comparison.
 - Shown in book using `anova()` command (p. 158)
- How does this work?

Module 3B: Significance Testing

- Model comparison: Comparison of relative fit of *nested models* to the data.
 - For each predictor, compare the fit of the “full model” to one that lacks that predictor
 - These models are nested—the full model “contains” the simpler model lacking a single predictor.
 - If the full model has a better fit to the data, that means that adding the predictor improves the model’s fit to the data; the predictor makes a significant contribution to the model

Module 3B: Significance Testing

- Likelihood is one way of measuring fit of a model to the data
 - Likelihood: the probability that a model assigns to some data (given the model, what's the probability of the data?)
 - This is what p-values provide; they are the likelihood of the data given the null hypothesis
- In the case of model comparison, given that the data *was* observed, we'd like to choose the model that assigns the data the highest likelihood—that is the model that has the best fit.

Module 3B: Significance Testing

- However, more complex models will tend, by chance, to have a better fit to the data
 - More free parameters = less constrained model → allows a better fit to the data
 - How do we tell when a more complex model is doing sufficient work to justify the extra complexity?

Module 3B: Significance Testing

- The likelihood ratio test is one means of testing this.
 - This test examines the ratio of the likelihood of the data under the simple model to the likelihood under the more complex model.
 - If this ratio is small enough, then we conclude that the more complex model is justified.
- The precise test statistic is a function of the *deviance*
 - $\text{deviance} = -2 \log \text{likelihood}$
 - $\text{Difference in deviances} = -2 \log \text{likelihood ratio}$

Module 3B: Significance Testing

Under certain (reasonable) assumptions, the null hypothesis that there's no difference in likelihood (beyond that predicted by an increase in the number of parameters) predicts that the difference in deviances should be distributed as χ^2

With degrees of freedom equal to the difference in the number of parameters across each model.

Here—since we are comparing a full model to one lacking a single parameter—the degrees of freedom should be 1

Take home point: The likelihood ratio test is a means of determining whether there's a significant improvement in model fit when a given parameter is included.

If $p < .05$ for this test, then the more complex model is justified—this parameter significantly improves model fit.

Module 3B: Significance Testing

- If you use model comparison (and you should!) you need to keep two things in mind.
 - Add the argument REML=F to your lmer command
 - The lmer algorithm uses restricted maximum likelihood to fit the parameters of your regression model.
 - This is not appropriate when using the likelihood ratio test.
 - You need to re-code all of your alphanumeric predictors into numerical codes within the dataframe that specifies your data.
 - We'll review how to do this after a brief example.

Module 3B: Significance Testing

- Basic procedure: Subtract fixed effects, leaving random effects (more below) constant
 - `full.lmer = lmer(RT~Frequency+(1+Frequency|Subject),data=lexdec, REML=F)`
 - `frequencyEffectTest.lmer = lmer(RT~Frequency - Frequency + (1+Frequency|Subject),data=lexdec, REML=F)`
 - `anova(full.lmer,frequencyEffectTest.lmer)`
- Why use this cumbersome procedure?
 - Makes clear in model specification what effects you are eliminating; easier to keep track of what effect you're testing.
- Note: When subtracting interactions, make sure to use the `:` notation to eliminate the interactions alone.

Module 3B: Significance Testing

- Issue: what do you do with multi-level factors?
 - e.g., place of articulation (bilabial, alveolar, velar)
 - We encode these three levels using two factors
 - Treatment: each level vs. baseline
 - Orthogonal contrasts (e.g., lingual vs. bilabial; alveolar vs. velar)
- The simple subtraction syntax won't work here
 - If you subtract place, you eliminate both
- Issue: contrast coding
 - If a factor isn't treatment coded, lme4 estimates random slopes for each level of the factor (along with their correlation).
 - That is not the intent of treatment coding!

Module 3B: Significance Testing

- Solution: Add columns to your dataframe that numerically represent contrast, and add them to your model.
 - You have to do the “translation” of alphanumeric codes into the numbers that represent each contrast
- Basic idea
 - Contrasts specify a vector.
 - E.g., -0.5,0,+0.5 is a list of 3 numbers
 - Look up values in this vector based on the number of the level of the factor.
 - If the factors are alveolar, bilabial, velar, for bilabial you should look up the 2nd number
- After you have columns, add them independently to the model.

Module 3B: Significance Testing

- Illustration: VOT place of articulation data.
- Execute code chunk 19. This reads in the data set and contrast codes place of articulation (just like we did in Module 2).
- Execute code chunk 20. This performs the “translation” of contrast codes into numeric codes and adds these columns to our dataframe.
- Execute code chunk 21. This builds an lmer using our numerically coded columns, and prints out the summary
- Execute code chunk 22. This builds an lmer using our alphanumeric column. Notice it’s identical to the output of code chunk 22—showing our translation was successful!

Module 3B: Significance Testing

- Because each contrast has been added independently, you can now separately subtract them from the model.
 - You also only get one random slope for each contrast (not one for each level of each contrast)
- Code chunks 23 and 24 perform this analysis for the bilabial vs. lingual and alveolar vs. velar contrast (respectively).
 - The anova() comparisons reveal that there is a significant contribution from each contrast, controlling for random variation across participants (variation in terms of overall VOT and effect of contrasts) and words (variation in terms of overall VOT).

Module 3B: Significance Testing

- **BONUS!** Now that your categorical variables are coded as numbers, you can examine their intercorrelations—a critical part of building robust regression models (remember Module 1!).
- As shown in code chunk 25, the correlation between these two orthogonal contrast is very low (by design!)
- You can also examine the residuals of the model. This is shown in code chunk 26.
 - Pretty good, but not perfect.
 - Inspection of VOT (code chunk 27) shows that this is probably because the VOT distribution is slightly skewed.
 - I would not be overly concerned by this deviation.

INTERIM SUMMARY

- Mixed effects models: Extend multiple regressions by controlling for idiosyncratic contributions of random effects —factors in your data that are drawn from larger populations (participants, items).
- Significance can be assessed just as we did with multiple regressions (t-tests, wald z), but I recommend using model comparison (more conservative).
- Don't forget modules 1+2! LMERS can break just as easily as regular multiple regressions.