Course Business

- Two new datasets for class today:
  - CourseWeb: Course Documents ➔ Sample Data ➔ Week 4

- Still need help with **lmerTest**?
- Other fixed effect questions?
  - Effect size—in lecture slides on CourseWeb. Will circle back to if we have time later

- Next two weeks: Random effects for different types of designs
  - This week: “Nested” random effects
  - Next week: “Crossed” random effects
Distributed Practice!

• What (if any) is the difference between this pair of models?
  • `lmer(StutteringFrequency ~ 1 + StutteringSeverity + QualityOfLife + (1|Subject) + (1|Item), data=stuttering)`
  • `lmer(StutteringFrequency ~ 1 + StutteringSeverity * QualityOfLife + (1|Subject) + (1|Item), data=stuttering)`

• What about this pair?
  • `lmer(WorkingMemory ~ 1 + Age * PhysicalActivity + (1|Subject), data=cog.aging)`
  • `lmer(WorkingMemory ~ 1 + Age + PhysicalActivity + Age:PhysicalActivity + (1|Subject), data=cog.aging)`
Distributed Practice!

• What (if any) is the difference between this pair of models?
  • `lmer(StutteringFrequency ~ 1 + StutteringSeverity + QualityOfLife + (1|Subject) + (1|Item), data=stuttering)`
  • `lmer(StutteringFrequency ~ 1 + StutteringSeverity * QualityOfLife + (1|Subject) + (1|Item), data=stuttering)`
  • The second model incorporates an *interaction* between stuttering severity and quality of life (in addition to the main effects)

• What about this pair?
  • `lmer(WorkingMemory ~ 1 + Age * PhysicalActivity + (1|Subject), data=cog.aging)`
  • `lmer(WorkingMemory ~ 1 + Age + PhysicalActivity + Age:PhysicalActivity + (1|Subject), data=cog.aging)`
  • NONE. These are just 2 different ways of writing the same model!
Week 4: Nested Random Effects

- Model Comparison
  - Nested Models
    - Hypothesis Testing
    - REML vs ML
  - Non-Nested Models
  - Shrinkage
- Nested Random Effects
  - Introduction to Clustering
  - Random Effects
  - Modeling Random Effects
  - Notation
  - Level-2 Variables
  - Multiple Random Effects
  - Limitations & Future Directions
Dataset

- Social support & health (e.g., Cohen & Wills, 1985)

- `lifeexpectancy.csv`:
  - Longitudinal study of 1000 subjects – some siblings from same family, so 517 total families
  - Perceived social support (z-scored)
  - Lifespan
  - And several control variables
Model Comparison

- Last week, we saw you could fit several different models from the same dataset
  - `model1 <- lmer(RT ~ 1 + PrevTrials + FontSize + (1|Subject) + (1|Item), data=Stroop)`
  - `model2 <- lmer(RT ~ 1 + PrevTrials + FontSize + PrevTrials:FontSize + (1|Subject) + (1|Item), data=Stroop)`
- Or:
  - `my.model <- lmer(Lifespan ~ 1 + SocSupport + YrsEducation + (1|Family), data=lifeexpectancy)`
  - `your.model <- lmer(Lifespan ~ 1 + HrsExercise + Conscientiousness + (1|Family), data=lifeexpectancy)`
Model Comparison

• One reason to save the results from each model is so that we can *compare* models:
  • Which model makes better predictions?
  • Compare which theoretical model better accounts for the data:
    ➢ Theoretical Model #1: Social support *does* affect health
    ➢ Theoretical Model #2: Social support *does not* affect health
**Nested Models**

- Three possible models of life expectancy:
  - Amount of weekly exercise
  - Amount of weekly exercise & perceived social support
  - Amount of weekly exercise, perceived social support, years of education, conscientiousness, yearly income, and number of vowels in your last name

- These are *nested* models—each one can be formed by subtracting variables from the one below it (“nested inside it”)
Nested Models

- Three possible models of life expectancy:
  - Amount of weekly exercise
  - Amount of weekly exercise & perceived social support
  - Amount of weekly exercise, perceived social support, years of education, conscientiousness, yearly income, and number of vowels in your last name

- Which set of information would give us the most accurate fitted() values?
Nested Models

- Three possible models of life expectancy:
  - Amount of weekly exercise
  - Amount of weekly exercise & perceived social support
  - Amount of weekly exercise, perceived social support, years of education, conscientiousness, yearly income, and number of vowels in your last name

- The “biggest” nested model will always provide predictions that are at least as good
- Adding info can only explain *more* of the variance
Nested Models

The “biggest” nested model will always provide predictions that are at least as good.
- Adding info can only explain more of the variance.
- Might not be much better (“number of vowels” effect zero or close to zero) but can’t be worse.

Slope of regression line relating last name vowels to life expectancy is near 0.
But that merely fails to improve predictions; doesn’t hurt them.

<table>
<thead>
<tr>
<th>Fixed effects:</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>54.68166</td>
<td>1.62613</td>
<td>33.63</td>
</tr>
<tr>
<td>HrsExercise</td>
<td>1.33084</td>
<td>0.11946</td>
<td>11.14</td>
</tr>
<tr>
<td>SocSupport</td>
<td>0.73018</td>
<td>0.24536</td>
<td>2.98</td>
</tr>
<tr>
<td>YrsEducation</td>
<td>0.56543</td>
<td>0.06385</td>
<td>8.86</td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>0.06708</td>
<td>0.23603</td>
<td>0.28</td>
</tr>
<tr>
<td>IncomeThousands</td>
<td>0.25837</td>
<td>0.02465</td>
<td>10.48</td>
</tr>
<tr>
<td>LastNameVowels</td>
<td>-0.06519</td>
<td>0.13437</td>
<td>-0.49</td>
</tr>
</tbody>
</table>
Likelihood Ratio Test

- We can compare nested models (only) using the likelihood-ratio test.
  - Remember that likelihood is what we search for in fitting an individual model (find the values with the highest likelihood).

- Likelihood is like the reverse of probability. Probability is about a result given a model. Likelihood is about a model given the results.
  - “Given a fair coin, what’s the probability of heads?” vs. “I got heads 83 out of 100 times. How likely is this to be a fair coin?”
Likelihood Ratio Test

- We can compare nested models (only) using the likelihood-ratio test.

- First, fit each of the models to be compared:
  - Try fitting a \texttt{model1} that includes both \texttt{HrsExercise} and \texttt{SocSupport} (with Family as a random effect).
  - Then, a \texttt{model2} that omits \texttt{SocSupport}. 

Likelihood Ratio Test

• Then, compare them with `anova()`:  
  • `anova(model1, model2)`  
  • Order doesn’t matter

Differences in log likelihoods are distributed as a **chi-square**  
• d.f. = number of variables we added or removed  
• Here, $\chi^2_{(1)} = 8.67, p < .01$
Week 4: Nested Random Effects

- Model Comparison
  - Nested Models
    - Hypothesis Testing
    - REML vs ML
  - Non-Nested Models
  - Shrinkage
- Nested Random Effects
  - Introduction to Clustering
  - Random Effects
  - Modeling Random Effects
  - Notation
  - Level-2 Variables
  - Multiple Random Effects
  - Limitations & Future Directions
Hypothesis Testing

Let’s think about our two models:

\[ E(Y_{i(jk)}) = \gamma_{000} + \gamma_{100} \text{HrsExercise} + \gamma_{200} \text{SocSupport} \]

\[ E(Y_{i(jk)}) = \gamma_{000} + \gamma_{100} \text{HrsExercise} \]

What are some possible values of \( \gamma_{200} \) (the SocSupport effect) in model 1?

- 3.83
- -1.04
- 0 – there is no social support effect
Hypothesis Testing

Let’s think about our two models:

\[ E(Y_{i(jk)}) = \gamma_{000} + \gamma_{100}\text{HrsExercise} + \gamma_{200}\text{SocSupport} \]

\[ E(Y_{i(jk)}) = \gamma_{000} + \gamma_{100}\text{HrsExercise} \]

What happens when \( \gamma_{200} \) is equal to 0?

- Anything multiplied by 0 is 0, so SocSupport just drops out of the equation
- Becomes the same thing as \textit{model2}
Hypothesis Testing

Let’s think about our two models:

\[ E(Y_{i(jk)}) = \gamma_{000} + \gamma_{100} \text{HrsExercise} + \gamma_{200} \text{SocSupport} \]

\[ E(Y_{i(jk)}) = \gamma_{000} + \gamma_{100} \text{HrsExercise} \]

\[ E(Y_{i(jk)}) = \gamma_{000} + \gamma_{100} \text{HrsExercise} + 0 \text{SocSupport} \]

- **model2** is just a **special case** of **model1**
  - The version of **model1** where \( \gamma_{200} = 0 \)
  - One of many possible versions of **model1**
  - Why we say **model2** is “nested” in **model1**
Hypothesis Testing

- Let’s think about our two models:

\[
\text{model 1: } E(Y_{i(jk)}) = \gamma_{000} + \gamma_{100}\text{HrsExercise} + \gamma_{200}\text{SocSupport}
\]

\[
\text{model 2: } E(Y_{i(jk)}) = \gamma_{000} + \gamma_{100}\text{HrsExercise}
\]

- This also helps show why \textit{model 1} always fits as well as \textit{model 2} or better
  - \textit{model 1} can account for the case where \(\gamma_{200} = 0\)
  - But it can also account for many other cases, too
Hypothesis Testing

Let’s think about our two models:

\[
\text{model1 } E(Y_{i(jk)}) = \gamma_{000} + \gamma_{100}\text{HrsExercise} + \gamma_{200}\text{SocSupport}
\]

\[
\text{model2 } E(Y_{i(jk)}) = \gamma_{000} + \gamma_{100}\text{HrsExercise} + \gamma_{200}\text{SocSupport}
\]

- Testing whether \textit{model2} fits significantly better \textit{is the same thing} as testing whether the SocSupport effect significantly differs from 0
  - i.e., whether there is a significant effect of SocSupport
- \textbf{LR test is another way of doing hypothesis testing!}
Hypothesis Testing

- Let’s think about our two models:

\[
\text{model 1 } \quad E(Y_{i(jk)}) = \gamma_{000} + \gamma_{100}\text{HrsExercise} + \gamma_{200}\text{SocSupport}
\]

\[
\text{E}(Y_{i(jk)}) = \gamma_{000} + \gamma_{100}\text{HrsExercise} + \gamma_{200}\text{SocSupport}
\]

\[
\text{model 2 } \quad E(Y_{i(jk)}) = \gamma_{000} + \gamma_{100}\text{HrsExercise}
\]

- “But you’re just comparing two models! You’re not actually testing the effect of social support!”

- Closely related to our research goal: Which theoretical model best explains data?
  - The theoretical model where social support doesn’t affect life expectancy
  - The model where social support does affect life expectancy
Model Comparison & Hypothesis Testing

- Ultimately, t-test and LR test very similar
  - **t-test**: Tests whether an effect differs from 0, based on this model
  - **Likelihood ratio**: Compare to a model where the effect actually IS constrained to be 0

```
p-value from lmerTest t-test: .0033
```

```
p-value from likelihood ratio test: .0032
```
Model Comparison & Hypothesis Testing

- Ultimately, t-test and LR test very similar
  - **t-test**: Tests whether an effect differs from 0, based on this model
  - **Likelihood ratio**: Compare to a model where the effect actually IS constrained to be 0

- In fact, with an infinitely large sample, these two tests would produce **identical** conclusions

- With small sample, **likelihood ratio** test is less likely to detect spurious differences
  - But, large differences uncommon
Week 4: Nested Random Effects

- Model Comparison
  - Nested Models
  - Hypothesis Testing
    - REML vs ML
  - Non-Nested Models
  - Shrinkage

- Nested Random Effects
  - Introduction to Clustering
  - Random Effects
  - Modeling Random Effects
  - Notation
  - Level-2 Variables
  - Multiple Random Effects
  - Limitations & Future Directions
REML vs ML

• Technically, *two* different algorithms that R can use “behind the scenes” to get the estimates

➤ **REML**: Restricted Maximum Likelihood
  • Assumes the fixed effects structure is correct
  • *Bad* for comparing models that differ in fixed effects

➤ **ML**: Maximum Likelihood
  • *OK* for comparing models
  • But, may underestimate variance of random effects

• Ideal: ML for model comparison, REML for final results
  • lme4 does this automatically for you!
  • Defaults to REML. But automatically refits models with ML when you do likelihood ratio test.
**REML vs ML**

- The one time you might have to mess with this:
- If you are going to be doing a lot of model comparisons, can fit the model with ML to begin with:
  - `model1 <- lmer(DV ~ Predictors, data=lifeexpectancy, REML=FALSE)`
  - Saves refitting for each comparison
  - Remember to refit the model with `REML=TRUE` for your final results
Week 4: Nested Random Effects

- Model Comparison
  - Nested Models
  - Hypothesis Testing
  - REML vs ML
- Non-Nested Models
- Shrinkage

- Nested Random Effects
  - Introduction to Clustering
  - Random Effects
  - Modeling Random Effects
  - Notation
  - Level-2 Variables
  - Multiple Random Effects
  - Limitations & Future Directions
Non-Nested Models

• Which of these pairs are cases of one model nested inside another? Which are not?
  • A
    • Accuracy ~ SentenceType + Aphasia + SentenceType:Aphasia
    • Accuracy ~ SentenceType + Aphasia
  • B
    • MathAchievement ~ SocioeconomicStatus
    • MathAchievement ~ TeacherRating + ClassSize
  • C
    • Recall ~ StudyTime
    • Recall ~ StudyTime + StudyStrategy
Non-Nested Models

• Which of these pairs are cases of one model nested inside another? Which are not?

• A
  • Accuracy ~ SentenceType + Aphasia + SentenceType:Aphasia
  • Accuracy ~ SentenceType + Aphasia

• B
  • MathAchievement ~ SocioeconomicStatus
  • MathAchievement ~ TeacherRating + ClassSize

• Each of these models has something that the other doesn’t have.
Non-Nested Models

- Models that aren’t nested can’t be tested the same way
- Nested model comparison was:

\[
E(Y_{i(jk)}) = \gamma_{000} + \gamma_{100} \text{HrsExercise} + \gamma_{200} \text{SocSupport}
\]

\[
E(Y_{i(jk)}) = \gamma_{000} + \gamma_{100} \text{HrsExercise} + 0
\]

- Null hypothesis \((H_0)\) is that there’s no SocSupport effect in the population (population parameter = 0)
- Could compare the observed SocSupport effect in our sample to the one we expect under \(H_0\) (0)
Non-Nested Models

• Models that aren’t nested can’t be tested the same way
• A non-nested comparison:

\[ E(Y_{i(jk)}) = Y_{000} + 0 \times \text{YrsEducation} + Y_{200} \times \text{IncomeThousands} \]

\[ E(Y_{i(jk)}) = Y_{000} + Y_{100} \times \text{YrsEducation} + 0 \times \text{IncomeThousands} \]

• What would support 1st model over 2nd?
  • \( Y_{200} \) is significantly greater than 0, but also \( Y_{100} \) is 0
  • But remember we can’t test that something is 0 with frequentist statistics … can’t prove the \( H_0 \) is true
  • Parametric statistics don’t apply here 😞
Non-Nested Models: Comparison

- **AIC**: An Information Criterion or Akaike’s Information Criterion
  - \(-2(\text{log likelihood}) + 2k\)
  - \(k = \# \text{ of fixed and random effects in a particular model}\)
  - A model with a lower AIC is better
  - Doesn’t assume any of the models is correct
  - Appropriate for **correlational / non-experimental** data

- **BIC**: Bayesian Information Criterion
  - \(-2(\text{log likelihood}) + \log(n)k\)
  - \(k = \# \text{ of fixed & random effects, } n = \text{num. observations}\)
  - A model with a lower BIC is better
  - Assumes that there’s a “true” underlying model in the set of variables being considered
  - Appropriate for **experimental** data

Yang, 2005; Oehlert, 2012
Non-Nested Models

- Can also get these from `anova()`
- Just **ignore** the chi-square if non-nested models

```r
> anova(model4,model5)
refitting model(s) with ML (instead of REML)
Data: lifeexpectancy
Models:
object: Lifespan ~ 1 + IncomeThousands + (1 | Family)
..1: Lifespan ~ 1 + YrsEducation + (1 | Family)

Df   AIC    BIC logLik deviance Chisq Chi Df Pr(>Chisq)
object 4 7068.7 7088.4  -3530.4  7060.7
..1   4 7090.5 7110.1  -3541.2  7082.5  0   0     0.0
```

- AIC and BIC do **not** have a significance test associated with them
- The model with the lower AIC/BIC is preferred, but we don’t know how *reliable* this preference is
Week 4: Nested Random Effects

- Model Comparison
  - Nested Models
  - Hypothesis Testing
  - REML vs ML
  - Non-Nested Models
- Shrinkage
- Nested Random Effects
  - Introduction to Clustering
  - Random Effects
  - Modeling Random Effects
  - Notation
  - Level-2 Variables
  - Multiple Random Effects
  - Limitations & Future Directions
Shrinkage

I was in the pool!!
**Shrinkage**

- Let’s try to predict your final grades in the class

<table>
<thead>
<tr>
<th>Paper 1</th>
<th>DESERVED SCORE</th>
<th>DARTBOARD OF SAMPLING ERROR!</th>
<th>RESULTING GRADE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length: 6 pages</td>
<td>90</td>
<td>10</td>
<td>100</td>
</tr>
<tr>
<td>Models run: 2</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Paper 2</th>
<th>DESERVED SCORE</th>
<th>DARTBOARD OF SAMPLING ERROR!</th>
<th>RESULTING GRADE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length: 5 pages</td>
<td>90</td>
<td>3</td>
<td>93</td>
</tr>
<tr>
<td>Models run: 1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Paper 3</th>
<th>DESERVED SCORE</th>
<th>DARTBOARD OF SAMPLING ERROR!</th>
<th>RESULTING GRADE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length: 3 pages</td>
<td>80</td>
<td>0</td>
<td>80</td>
</tr>
<tr>
<td>Models run: 4</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Shrinkage

• Page length seems like a good predictor of grades, but partially due to sampling error.
• All parameter estimates influenced by noise in data.

<table>
<thead>
<tr>
<th>Paper</th>
<th>Length</th>
<th>Models run</th>
<th>DESERVED SCORE</th>
<th>DARTBOARD OF SAMPLING ERROR</th>
<th>RESULTING GRADE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paper 1</td>
<td>6 pages</td>
<td>2</td>
<td>90</td>
<td>10</td>
<td>100</td>
</tr>
<tr>
<td>Paper 2</td>
<td>5 pages</td>
<td>1</td>
<td>90</td>
<td>3</td>
<td>93</td>
</tr>
<tr>
<td>Paper 3</td>
<td>3 pages</td>
<td>4</td>
<td>80</td>
<td>0</td>
<td>80</td>
</tr>
</tbody>
</table>
Shrinkage

• Our estimates (and any choice of variables resulting from this) always partially reflect the idiosyncrasies/noise in the data set we used to obtain them

• Won’t fit any later data set quite as well … shrinkage

• Problem when we’re using the data to decide the model
• In experimental context, design/model usually known in advance
**Shrinkage**

- Our estimates (and any choice of variables resulting from this) always partially reflect the idiosyncrasies/noise in the data set we used to obtain them.

- Won’t fit any later data set quite as well ... **shrinkage**

- “*If you use a sample to construct a model, or to choose a hypothesis to test, you cannot make a rigorous scientific test of the model or the hypothesis using that same sample data.*” (Babyak, 2004, p. 414)
Shrinkage—Examples

• Adak Island, Alaska
  • Daily temperature here predicts stock market activity!
  • \( r = -.87 \) correlation with the price of a specific group of stocks!
  • Completely true—I’m not making this up!

• Problem with this:
  • With thousands of weather stations & stocks, easy to find a strong correlation somewhere, even if it’s just sampling error
  • Problem is that this factoid doesn’t reveal all of the other (non-significant) weather stations & stocks we searched through
  • Would only be impressive if this hypothesis continued to be true on a new set of weather data & stock prices

Vul et al., 2009
Shrinkage—Examples

• “Voodoo correlations” issue in some fMRI analyses (Vul et al., 2009)
  • Find just the voxels (parts of a brain scan) that correlate with some outcome measure (e.g., personality)

• Then, report the average activation in those voxels with the personality measure

• Voxels were already chosen on the basis of those high correlations
  • Thus, includes sampling error favoring the correlation but excludes error that doesn’t
  • Real question is whether the chosen voxels would predict personality in a new, independent dataset
Shrinkage—Solutions

- We need to be careful when using the data to select between models.

- The simplest solution: Test if a model obtained from one subset of the data applies to another subset (validation)
  - e.g., training and test sets

- The better solution: Do this with many randomly chosen subsets
  - Monte Carlo methods
  - Reading on CourseWeb for some general ways to do this in R
Week 4: Nested Random Effects

- Model Comparison
  - Nested Models
  - Hypothesis Testing
  - REML vs ML
- Non-Nested Models
- Shrinkage

- Nested Random Effects
  - Introduction to Clustering
  - Random Effects
  - Modeling Random Effects
  - Notation
  - Level-2 Variables
  - Multiple Random Effects
  - Limitations & Future Directions
Theories of Intelligence

Which is closer to your view?

- You have a certain amount of intelligence, and you can’t really do much to change it.

  FIXED MINDSET

- Intelligence can be developed and is something you have to work for.

  GROWTH MINDSET
Theories of Intelligence

• Growth mindset has been linked to greater persistence & success in academic (& other work) (Dweck, 2008)

• Let’s see if this is true for middle-schoolers’ math achievement
  • math.csv on CourseWeb (Sample Data, Week 4)
  • 30 students in each of 24 classrooms (N = 720)
  • Measure growth mindset … 0 to 7 questionnaire

  “You can learn new things, but you can’t really change your basic intelligence.”

• Dependent measure: Score on an end-of-year standardized math exam (0 to 100)
Theories of Intelligence

- We can start writing a regression line to relate growth mindset to end-of-year score

\[ Y_{i(j)} = Y_{100} X_{1i(j)} \]

End-of-year math exam score

Growth mindset
Theories of Intelligence

- What about kids whose Growth Mindset score is 0?
  - Completely Fixed mindset
  - Even these kids probably will score at least some points on the math exam; won’t completely bomb
- Include an intercept term
  - Math score when theory of intelligence score = 0

\[ Y_{i(j)} = Y_{000} + Y_{100}x_{1i(j)} \]

End-of-year math exam score
Baseline
Growth mindset
Theories of Intelligence

- We probably can’t predict each student’s math score exactly.
- Kids differ in ways other than their growth mindset.
- Include an error term.
- Residual difference between predicted & observed score for observation \( i \) in classroom \( j \).
- Captures what’s unique about child \( i \).
- Assume these are independently, identically normally distributed (mean 0).

\[
Y_{i(j)} = Y_{000} + Y_{100} \times x_{1i(j)} + E_{i(j)}
\]

- \( Y_{i(j)} \): End-of-year math exam score.
- \( Y_{000} \): Baseline.
- \( Y_{100} \times x_{1i(j)} \): Growth mindset.
- \( E_{i(j)} \): Error.
Theories of Intelligence Data

Sampled CLASSROOMS
- Ms. Fulton’s Class
- Mr. Wagner’s Class
- Ms. Green’s Class
- Ms. Cornell’s Class

Sampled STUDENTS
- Student 1
- Student 2
- Student 3
- Student 4

Math achievement score $y_{11}$
Theory of intelligence score $x_{111}$
Independent error term $e_{11}$

Math achievement score $y_{21}$
Theory of intelligence score $x_{121}$
Independent error term $e_{21}$

Math achievement score $y_{42}$
Theory of intelligence score $x_{142}$
Independent error term $e_{42}$

- Where is the problem here?
Theories of Intelligence Data

Sampled CLASSROOMS
- Differences in classroom size, teaching style, teacher’s experience…

Sampled STUDENTS
- Students in the same classroom probably have more similar scores. **Clustering.**

- Error terms not fully independent
**Clustering**

- Why does clustering matter?
- Remember that we test effects by comparing them to their standard error:

\[ t = \frac{\text{Estimate}}{\text{Std. error}} \]

\( t \) = Estimate

Std. error

But if we have a lot of kids from the same classroom, they share more similarities than all kids in population

*Understating* the standard error across subjects…

…thus *overstating* the significance test

- Failing to account for clustering can lead us to detect spurious results (sometimes quite badly!)
Week 4: Nested Random Effects

- Model Comparison
  - Nested Models
  - Hypothesis Testing
  - REML vs ML
- Non-Nested Models
- Shrinkage
- Nested Random Effects
  - Introduction to Clustering
  - Random Effects
  - Modeling Random Effects
  - Notation
  - Level-2 Variables
  - Multiple Random Effects
  - Limitations & Future Directions
Random Effects

- Can’t we just add Classroom as another fixed effect variable?
- $1 + \text{TOI} + \text{Classroom}$
- Not what we want for several reasons
  - e.g., We’d get many, many comparisons between individual classrooms

| Coefficients: | Estimate | Std. Error | t value | $Pr(>|t|)$ |
|---------------|----------|------------|---------|-----------|
| (Intercept)   | 76.07314 | 1.06108    | 71.694  | <2e-16 ***|
| TOI           | 1.46009  | 0.08983    | 16.254  | <2e-16 ***|
| ClassroomBrehm| -9.03434 | 1.42491    | -6.340  | 4.13e-10 ***|
| ClassroomCarol| -2.45822 | 1.42241    | -1.728  | 0.084396 |
| ClassroomChang| -7.06268 | 1.42231    | -4.966  | 8.62e-07 ***|
| ClassroomChavez| -4.53531 | 1.42226    | -3.189  | 0.001493 **|
| ClassroomCornell| -3.84651 | 1.42352    | -2.702  | 0.007058 **|
| ClassroomCreighton| -10.31291| 1.42352    | -7.245  | 1.15e-12 ***|
| ClassroomEnschede| -9.68123 | 1.42237    | -6.806  | 2.16e-11 ***|
| ClassroomFranklin| -1.45993 | 1.42226    | -1.026  | 0.305018 |
| ClassroomFulton| -5.11620 | 1.42317    | -3.595  | 0.000347 ***|
| ClassroomGreen| -5.68466 | 1.42226    | -3.997  | 7.10e-05 ***|
| ClassroomHartmann| -7.97709 | 1.42251    | -5.608  | 2.96e-08 ***|
| ClassroomHouston| -10.17496| 1.42251    | -7.153  | 2.16e-12 ***|
| ClassroomJackson| -3.47281 | 1.42257    | -2.441  | 0.014886 *|
| ClassroomLam| -6.15548 | 1.42264    | -4.327  | 1.73e-05 ***|
| ClassroomMatsumoto| -2.72472 | 1.42257    | -1.915  | 0.055858 .|
| ClassroomRyskin| -0.03451 | 1.42340    | -0.024  | 0.980664 |
| ClassroomSmith| -5.69015 | 1.42264    | -4.000  | 7.02e-05 ***|
| ClassroomStoddard| -3.10356 | 1.42317    | -2.181  | 0.029537 *|
| ClassroomStuart| -4.29443 | 1.42241    | -3.019  | 0.002628 **|
| ClassroomTrude| -4.20310 | 1.42257    | -2.955  | 0.003237 **|
| ClassroomTullis| -3.61104 | 1.42246    | -2.539  | 0.011347 *|
| ClassroomWagner| -8.26124 | 1.42306    | -5.805  | 9.77e-09 ***|
| ClassroomWilliams| -2.07705 | 1.42234    | -1.460  | 0.144657 |
Random Effects

• What makes the Classroom variable different from the TOI variable?
  ➢ Theoretical interest is in effects of theories of intelligence, not in effects of being Ms. Fulton
  ➢ If another researcher wanted to replicate this experiment, they could include the Theories of Intelligence scale, but they probably couldn’t get the same teachers
  ➢ We do expect our results to generalize to other teachers/classrooms, but this experiment doesn’t tell us anything about how the relation would generalize to other questionnaires

• These classrooms are just some classrooms we sampled out of the population of interest
Fixed Effects vs. Random Effects

**Fixed effects:**
- We’re interested in the specific categories/levels
- The categories are a complete set
  - At least within the context of the experiment

**Random effects:**
- Not interested in the specific categories
- For a *true* “random effect“:
  - The observed set of categories samples from a larger population
  - The population is large enough that we can estimate its variance (e.g., 100 times # of sampled levels)
  - Might call this a **variable intercept model** (+ variable slopes, later) if we’re not trying to generalize to a population
Random Effect or Fixed Effect?

• Scott interested in the effects of distributed practice on grad students’ statistics learning. For his experimental items, he picks 10 statistics formulae randomly out of a textbook. Then, he samples 20 Pittsburgh-area grad students as participants. Half study the items using distributed practice and half study using massed practice (a single day) before they are all tested.
• **Participant** is a...

• **Item** is a...

• **Practice type** (distributed vs. massed) is a ...
Random Effect or Fixed Effect?

- Scott interested in the effects of distributed practice on grad students’ statistics learning. For his experimental items, he picks 10 statistics formulae randomly out of a textbook. Then, he samples 20 Pittsburgh-area grad students as participants. Half study the items using distributed practice and half study using massed practice (a single day) before they are all tested.

- **Participant** is a…
  - Random effect. Scott sampled them out of a much larger population of interest (grad students).

- **Item** is a…
  - Random effect. Scott’s not interested in these specific formulae; he picked them out randomly.

- **Practice type** (distributed vs. massed) is a …
  - Fixed effect. We’re comparing these 2 specific conditions
Random Effect or Fixed Effect?

• A researcher in education is interested in the relation between class size and student evaluations at the university level. The research team collects data at 10 different universities across the US. **University** is a…

• A planner for the city of Pittsburgh compares the availability of parking at Pitt vs CMU. **University** is a…
Random Effect or Fixed Effect?

• A researcher in education is interested in the relation between class size and student evaluations at the university level. The research team collects data at 10 different universities across the US. **University** is a…
  • Random effect. Goal is to generalize to universities as a whole, and we just sampled these 10.

• A planner for the city of Pittsburgh compares the availability of parking at Pitt vs CMU. **University** is a…
  • Fixed effect. Now, we DO care about these two particular universities.
Random Effect or Fixed Effect?

• We’re studying students learning to speak English as a second language. Our goal is to compare their productions of regular vs irregular verbs. However, we also need to account for the fact that our participant speak a variety of different first languages, which is a…
Random Effect or Fixed Effect?

• We’re studying students learning to speak English as a second language. Our goal is to compare their productions of regular vs irregular verbs. However, we also need to account for the fact that our participants speak a variety of different first languages, which is a…

• Random effect. We’re not interested in specific languages, and the languages represented by our sample are probably only a set of all possible first languages.
Random Effect or Fixed Effect?

• We’re interested in how inferred causality between two variables is affected by the strength of relation. In our experimental design, we manipulate the correlation to be either 0, .25, or .75. **Strength of relation** is a…
Random Effect or Fixed Effect?

• We’re interested in how inferred causality between two variables is affected by the strength of relation. In our experimental design, we manipulate the correlation to be either 0, .25, or .75. Strength of relation is a…

• Fixed effect. This is the variable that we’re theoretically interested in and want to model. Also, 0, .25, and .75 exhaustively characterize strength of relation within this experimental design.
Week 4: Nested Random Effects

- Model Comparison
  - Nested Models
    - Hypothesis Testing
    - REML vs ML
  - Non-Nested Models
  - Shrinkage

- Nested Random Effects
  - Introduction to Clustering
  - Random Effects
  - Modeling Random Effects
  - Notation
  - Level-2 Variables
  - Multiple Random Effects
  - Limitations & Future Directions
Modeling Random Effects

• Let’s add Classroom as a random effect to the model (then we’ll talk about what it’s doing)

• Can you fill in the rest?
• `model1 <- lmer(FinalMathScore ~ 1 + TOI + (1|Classroom), data=math)`
Modeling Random Effects

• Let’s add Classroom as a random effect to the model (then we’ll talk about what it’s doing)

• Can you fill in the rest?
• `model1 <- lmer(FinalMathScore ~ 1 + TOI + (1|Classroom), data=math)`
Modeling Random Effects

• Let’s add Classroom as a random effect to the model (then we’ll talk about what it’s doing)

• Can you fill in the rest?
  • `model1 <- lmer(FinalMathScore ~ 1 + TOI + (1 | Classroom), data=math)`

• We’re allowing each classroom to have a different intercept
  • Some classrooms have higher math scores on average
  • Some have lower math scores on average
  • A random intercept
Modeling Random Effects

• Let’s add Classroom as a random effect to the model (then we’ll talk about what it’s doing)

• Can you fill in the rest?
  • `model1 <- lmer(FinalMathScore ~ 1 + TOI + (1|Classroom), data=math)`

• We are not interested in comparing the specific classrooms we sampled
• Instead, we are model the *variance* of this population
• How much do classrooms typically vary in math achievement?
Modeling Random Effects

- Model results:

<table>
<thead>
<tr>
<th>Groups</th>
<th>Name</th>
<th>Variance</th>
<th>Std.Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classroom (Intercept)</td>
<td></td>
<td>8.198</td>
<td>2.863</td>
</tr>
<tr>
<td>Residual</td>
<td></td>
<td>30.343</td>
<td>5.508</td>
</tr>
</tbody>
</table>

- We are not interested in comparing the specific classrooms we sampled.
- Instead, we are modeling the variance of this population.
- How much do classrooms typically vary in math achievement?
Week 4: Nested Random Effects

- Model Comparison
  - Nested Models
  - Hypothesis Testing
  - REML vs ML
  - Non-Nested Models
  - Shrinkage

- Nested Random Effects
  - Introduction to Clustering
  - Random Effects
  - Modeling Random Effects
  - Notation
  - Level-2 Variables
  - Multiple Random Effects
  - Limitations & Future Directions
Notation

• What exactly is this model doing?

• Let’s go back to our model of individual students (now slightly different):
Notation

• What exactly is this model doing?

What now determines the baseline that we should expect for students with growth mindset=0?

• Let’s go back to our model of individual students (now slightly different):

\[ Y_{i(j)} = B_{00j} + Y_{100}X_{1i(j)} + E_{i(j)} \]

- \( Y_{i(j)} \): End-of-year math exam score
- \( B_{00j} \): Baseline
- \( Y_{100} \): Growth mindset
- \( E_{i(j)} \): Student Error
Notation

• What exactly is this model doing?
• Baseline (intercept) for a student in classroom $j$ now depends on two things:

$B_{00j} = Y_{000} + U_{0j}$

- Intercept
- Overall intercept across everyone
- Teacher effect for this classroom (Error)

• Let’s go back to our model of individual students (now slightly different):

$Y_{i(j)} = B_{00j} + Y_{100}x_{1i(j)} + E_{i(j)}$

- End-of-year math exam score
- Baseline
- Growth mindset
- Student Error
Notation

- Essentially, we have two regression models
  - Hierarchical linear model
  - Model of classroom $j$:
    
    \[ B_{00j} = Y_{000} + U_{0j} \]

    - Intercept
    - Overall intercept across everyone
    - Teacher effect for this classroom (Error)

- Model of student $i$:
  
  \[ Y_{i(j)} = B_{00j} + Y_{100}x_{1i(j)} + E_{i(j)} \]

  - End-of-year math exam score
  - Baseline
  - Growth mindset
  - Student Error
Level-2 model is for the superordinate level here, Level-1 model is for the subordinate level.
Notation

- Two models seems confusing. But we can simplify with some algebra…
  - Model of classroom $j$:
    \[ B_{00j} = Y_{000} + U_{0j} \]
    - $B_{00j}$: Intercept
    - $Y_{000}$: Overall intercept across everyone
    - $U_{0j}$: Teacher effect for this classroom (Error)
  - Model of student $i$:
    \[ Y_{i(j)} = B_{00j} + Y_{100}x_{1i(j)} + E_{i(j)} \]
    - $Y_{i(j)}$: End-of-year math exam score
    - $B_{00j}$: Baseline
    - $Y_{100}x_{1i(j)}$: Growth mindset
    - $E_{i(j)}$: Student Error

LEVEL-2 MODEL (Classroom)
LEVEL-1 MODEL (Student)
Notation

- Substitution gives us a *single* model that combines level-1 and level-2
  - **Mixed effects model**
  - Combined model:

\[
Y_{i(j)} = Y_{000} + U_{0j} + Y_{100} X_{1i(j)} + E_{i(j)}
\]

- $Y_{i(j)}$: End-of-year math exam score
- $Y_{000}$: Overall intercept
- $U_{0j}$: Teacher effect for this classroom (Error)
- $Y_{100} X_{1i(j)}$: Growth mindset
- $E_{i(j)}$: Student Error
Notation

• Just two slightly different ways of writing the same thing. **Notation** difference, not statistical!
• Mixed effects model:

\[
Y_{i(j)} = Y_{000} + U_{0j} + Y_{100}x_{1i(j)} + E_{i(j)}
\]

• Hierarchical linear model:

\[
B_{00j} = Y_{000} + U_{0j}
\]

\[
Y_{i(j)} = B_{00j} + Y_{100}x_{1i(j)} + E_{i(j)}
\]
lme4 always uses the mixed-effects model notation

\[ Y_{i(j)} = Y_{000} + Y_{100}x_{1i(j)} + U_{0j} + E_{i(j)} \]

End-of-year math exam score

Overall intercept

Growth mindset

Teacher effect for this class (Error)

Student Error

lmer(
FinalMathScore ~ 1 + TOI + (1|Classroom)
)

(Level-1 error is always implied, don’t have to include)
Week 4: Nested Random Effects

- Model Comparison
  - Nested Models
    - Hypothesis Testing
    - REML vs ML
  - Non-Nested Models
  - Shrinkage

- Nested Random Effects
  - Introduction to Clustering
  - Random Effects
  - Modeling Random Effects
  - Notation
  - Level-2 Variables
  - Multiple Random Effects
  - Limitations & Future Directions
Level-2 Variables

- So far, all our model says about classrooms is that they’re different
  - Some classrooms have a large intercept
  - Some classrooms have a small intercept

- But, we might also have some interesting variables that characterize classrooms
  - They might even be our main research interest!

- How about teacher theories of intelligence?
  - Might affect how they interact with & teach students
Level-2 Variables

- **TeacherTheory** characterizes Level 2
- *All* students in the same classroom will have the *same* **TeacherTheory**
- `xtabs(~ TeacherTheory + Classroom, data=math)`
Level-2 Variables

- This becomes another variable in the level-2 model of classroom differences
- Tells us what we can expect this classroom to be like

\[
B_{00j} = Y_{000} + Y_{200}X_{20j} + U_{0j}
\]

\[
Y_{i(j)} = B_{00j} + Y_{100}X_{1i(j)} + E_{i(j)}
\]

- \(B_{00j}\): Intercept
- \(Y_{000}\): Overall intercept
- \(Y_{200}\): Teacher mindset
- \(U_{0j}\): Teacher effect for this classroom (Error)
- \(Y_{100}\): Growth mindset
- \(E_{i(j)}\): Student Error
- \(X_{1i(j)}\): Baseline

\(\text{LEVEL-2 MODEL (Classroom)}\)
\(\text{LEVEL-1 MODEL (Student)}\)
Level-2 Variables

- Teacher mindset is a fixed-effect variable
  - We ARE interested in the effects of teacher mindset on student math achievement … a research question, not just something to control for

- Even if we ran this with a new random sample of 30 teachers, we WOULD hope to replicate whatever regression slope for teacher mindset we observe (whereas we wouldn’t get the same 30 teachers back)
**Level-2 Variables**

- Since R uses mixed effects notation, we don’t have to do anything special to add a level-2 variable to the model.

  ```r
  model2 <- lmer(FinalMathScore ~ 1 + TOI + TeacherTheory + (1|Classroom), data=math)
  ```

- R automatically figures out `TeacherTheory` is a level-2 variable because it’s invariant for each classroom.
- We keep the random intercept for Classroom because we don’t expect `TeacherTheory` will explain all of the classroom differences. Intercept captures residual differences.
Week 4: Nested Random Effects

- Model Comparison
  - Nested Models
  - Hypothesis Testing
  - REML vs ML
  - Non-Nested Models
  - Shrinkage
- Nested Random Effects
  - Introduction to Clustering
  - Random Effects
  - Modeling Random Effects
  - Notation
  - Level 2 Variables
- Multiple Random Effects
- Limitations & Future Directions
Hold on! Classrooms aren’t fully independent, either. Some of them are from the same school, and some are from different schools.
Multiple Random Effects

Is SCHOOL a fixed effect or a random effect?
- These schools are just a sample of possible schools of interest -> Random effect.
Multiple Random Effects

LEVEL 3
Sampled SCHOOLS

LEVEL 2
Sampled CLASSROOMS

LEVEL 1
Sampled STUDENTS

- No problem to have more than 1 random effect in the model! Try adding a random intercept for school.
Multiple Random Effects

LEVEL 3
Sampled SCHOOLS

LEVEL 2
Sampled CLASSROOMS

LEVEL 1
Sampled STUDENTS

• `model3 <- lmer(FinalMathScore ~ 1 + TOI + TeacherTheory + (1|Classroom) + (1|School), data=math)`
Multiple Random Effects

LEVEL 3
Sampled SCHOOLS

LEVEL 2
Sampled CLASSROOMS

LEVEL 1
Sampled STUDENTS

- This is an example of **nested** random effects.
- Each classroom is always in the same school.
Multiple Random Effects

LEVEL 3
Sampled SCHOOLS

LEVEL 2
Sampled CLASSROOMS

LEVEL 1
Sampled STUDENTS

- Let’s do an intervention: Hours of use of math tutoring software
- Which level(s) of the model could this be at?
Multiple Random Effects

LEVEL 3
Sampled SCHOOLS
Carnegie Learning
If use of the tutor characterizes a whole school

LEVEL 2
Sampled CLASSROOMS

LEVEL 1
Sampled STUDENTS

• Let’s do an intervention: Hours of use of math tutoring software
• Which level(s) of the model could this be at?
Multiple Random Effects

Let’s do an intervention: Hours of use of math tutoring software

Which level(s) of the model could this be at?
Let's do an intervention: Hours of use of math tutoring software

Which level(s) of the model could this be at?
Multiple Random Effects

LEVEL 3
Sampled SCHOOLS

LEVEL 2
Sampled CLASSROOMS

LEVEL 1
Sampled STUDENTS

• Can you find this out from R?
Multiple Random Effects

Can you find this out from R?
• `xtabs(~TutorHours + Classroom, data=math)`
Try adding TutorHours to your model

```r
model4 <- lmer(FinalMathScore ~ 1 + TOI + TutorHours + TeacherTheory + (1|Classroom) + (1|School), data=math)
```

Don’t need to specify level; `lmer()` figures it out!

But, important for interpretation
Week 4: Nested Random Effects

- Model Comparison
  - Nested Models
    - Hypothesis Testing
    - REML vs ML
  - Non-Nested Models
  - Shrinkage
- Nested Random Effects
  - Introduction to Clustering
  - Random Effects
  - Modeling Random Effects
  - Notation
  - Level 2 Variables
  - Multiple Random Effects
  - Limitations & Future Directions


Limitations & Future Directions

• #1: Assuming classrooms differ only in intercept (overall math score)
• But slope of regression line for tutor use might vary across schools. Used more or less effectively
Limitations & Future Directions

LEVEL 3
Sampled SCHOOLS

LEVEL 2
Sampled CLASSROOMS

LEVEL 1
Sampled STUDENTS

• #2: Random effects here are fully nested
  • Each student in 1 classroom, each classroom in 1 school
Limitations & Future Directions

• #2: Random effects here are fully nested
• But what about something like this?
• Each subject seems more than 1 item, each item presented to more than 1 subject
Limitations & Future Directions

• We will address both of these next week 😊