Week 6.1: Crossed Random Effects

- Crossed Random Effects Analysis
  - Example Design
    - Joining Dataframes

- Random Slopes in Crossed Designs
  - Between-Subjects & Within-Subjects Designs
  - Between-Items & Within-Items Designs
  - When are Random Slopes Necessary?
  - Practice Activity
An Experimental Dataset

• naming.csv
  • RT (response time) to “name” a word (read it aloud) for students learning English
  • 60 Subjects each presented with 49 Words we randomly picked out of a dictionary
  • Each row of data is a single trial (one subject responding to one word)
  • We’re interested in YearsOfStudy (of English) and WordFreq (frequency in the English language)
    • Also interested in their interaction
• Which of these should we consider Fixed Effects? Which are Random Effects?
An Experimental Dataset

• naming.csv
  • RT (response time) to "name" a word (read it aloud) for students learning English
  • 60 Subjects each presented with 49 Words we randomly picked out of a dictionary
  • Each row of data is a single trial (one subject responding to one word)
  • We’re interested in YearsOfStudy (of English) and WordFreq (frequency in the English language)
    • Also interested in their interaction

• Which of these should we consider Fixed Effects? Which are Random Effects?
  • Fixed: YearsOfStudy, WordFreq, & interaction
  • Random: Subject, Word
Now, we have >1 trial per Subject.
- Level-1 observations (RTs) are nested within subjects.
- Clear we need to take account of this.
  - Some subjects will have faster RTs in general than others.
**Crossed Random Effects**

- Here are the different **words**. Do some seem easier than others?

<table>
<thead>
<tr>
<th>astronaut</th>
<th>boy</th>
<th>breakfast</th>
<th>carburetor</th>
<th>chair</th>
<th>cheese</th>
<th>chemistry</th>
<th>clogs</th>
<th>coffee</th>
</tr>
</thead>
<tbody>
<tr>
<td>60</td>
<td></td>
<td>60</td>
<td>60</td>
<td>60</td>
<td>60</td>
<td>60</td>
<td>60</td>
<td>60</td>
</tr>
<tr>
<td>computer</td>
<td>dentist</td>
<td>dolphin</td>
<td>door</td>
<td>drive</td>
<td>duck</td>
<td>eat</td>
<td>girl</td>
<td>glasses</td>
</tr>
<tr>
<td>60</td>
<td>60</td>
<td>60</td>
<td>60</td>
<td>60</td>
<td>60</td>
<td>60</td>
<td>60</td>
<td>60</td>
</tr>
<tr>
<td>green</td>
<td>hair</td>
<td>lawn</td>
<td>marigold</td>
<td>memorandum</td>
<td>monorail</td>
<td>orange</td>
<td>overturned</td>
<td>pajamas</td>
</tr>
<tr>
<td>60</td>
<td>60</td>
<td>60</td>
<td>60</td>
<td>60</td>
<td>60</td>
<td>60</td>
<td>60</td>
<td>60</td>
</tr>
<tr>
<td>panther</td>
<td>pedal</td>
<td>penguin</td>
<td>peony</td>
<td>perplex</td>
<td>pomegranate</td>
<td>potato</td>
<td>prenups</td>
<td>resend</td>
</tr>
<tr>
<td>60</td>
<td>60</td>
<td>60</td>
<td>60</td>
<td>60</td>
<td>60</td>
<td>60</td>
<td>60</td>
<td>60</td>
</tr>
<tr>
<td>sell</td>
<td>send</td>
<td>shower</td>
<td>slow</td>
<td>suggest</td>
<td>talk</td>
<td>teach</td>
<td>traffic</td>
<td>train</td>
</tr>
<tr>
<td>60</td>
<td>60</td>
<td>60</td>
<td>60</td>
<td>60</td>
<td>60</td>
<td>60</td>
<td>60</td>
<td>60</td>
</tr>
<tr>
<td>type</td>
<td>verify</td>
<td>walk</td>
<td>yell</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>60</td>
<td>60</td>
<td>60</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Crossed Random Effects

- Each word is also used in more than 1 trial
- Want to take account of that, too
  - Observations from the same item will be more similar to one another, too ("boy" easier than "carburetor")
Crossed Random Effects

In fact, we can think of each trial as the *pairing* of a subject and a word.
Crossed Random Effects

- Random effects not hierarchically nested
- Before: Each classroom appears in only 1 school
- Here: Each item presented to each subject
- If we draw all the lines in, we see they cross
Crossed Random Effects

Subjects (Level-2)

Trials (Level-1)

Words (Level-2)

- **Crossed random effects** structure
- *a/k/a* **cross-classified** when we’re dealing with existing classifications
Crossed Random Effects

• Conceptually different sampling, but same syntax!
• Let’s try fitting a model with:
  • A fixed effect of YearsOfStudy
  • Random intercepts for Subject and Word

```r
model1 <- lmer(RT ~ YearsOfStudy + (1|Subject) + (1|Word), data=naming)
```
Crossed Random Effects

- Conceptually different sampling, but same syntax!
- Let’s try fitting a model with:
  - A fixed effect of `Years0fStudy`
  - Random intercepts for `Subject` and `Word`

```r
model1 <- lmer(RT ~ 1 + Years0fStudy + (1|Subject) + (1|Word), data=naming)
```
Crossed Random Effects

- `model1 <- lmer(RT ~ 1 + YearsOfStudy + (1|Subject) + (1|Word), data=naming)`

Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']

Formula: RT ~ 1 + YearsOfStudy + (1 | Subject) + (1 | Word)

Data: naming

REML criterion at convergence: 34208.4

Scaled residuals:

<table>
<thead>
<tr>
<th></th>
<th>Min</th>
<th>1Q</th>
<th>Median</th>
<th>3Q</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-3.6458</td>
<td>-0.6466</td>
<td>-0.0037</td>
<td>0.6831</td>
<td>3.3835</td>
</tr>
</tbody>
</table>

Random effects:

<table>
<thead>
<tr>
<th>Groups</th>
<th>Name</th>
<th>Variance</th>
<th>Std.Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject</td>
<td>(Intercept)</td>
<td>2353</td>
<td>48.51</td>
</tr>
<tr>
<td>Word</td>
<td>(Intercept)</td>
<td>19091</td>
<td>138.17</td>
</tr>
<tr>
<td>Residual</td>
<td>(Intercept)</td>
<td>5727</td>
<td>75.68</td>
</tr>
</tbody>
</table>

Number of obs: 2940, groups: Subject, 60; Word, 49

Fixed effects:

|                  | Estimate | Std. Error | df | t value | Pr(>|t|) |
|------------------|----------|------------|----|---------|---------|
| (Intercept)      | 590.803  | 25.305     | 95.844 | 23.348  | < 2e-16 ** |
| YearsOfStudy     | -36.809  | 4.002      | 58.000 | -9.197  | 6.29e-13 *** |

---

Signif. codes:  < 0.001 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:

<table>
<thead>
<tr>
<th></th>
<th>(Intr)</th>
<th>YearsOfStdy</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intr)</td>
<td>-0.572</td>
<td></td>
</tr>
</tbody>
</table>

Significant effect of YearsOfStudy – more study = faster responding
Crossed Random Effects

Subjects (Level-2)
- Subject 1
- Subject 2

Trials (Level-1)
- RT 1
- RT 2
- RT 3
- RT 4

Words (Level-2)
- "Boy"
- "Carburator"

- Huge improvement over ANOVA analyses in psycholinguistics / experimental psychology!
- We used to have to conduct separate analyses to assess generalizability over subjects & items
- Or, item effects were simply ignored!
Week 6.1: Crossed Random Effects

- Crossed Random Effects Analysis
  - Example Design
  - Joining Dataframes

- Random Slopes in Crossed Designs
  - Between-Subjects & Within-Subjects Designs
  - Between-Items & Within-Items Designs
  - When are Random Slopes Necessary?
  - Practice Activity
Joining Dataframes

• Let’s also look at the effect of word frequency
• Problem: This is currently saved in a separate file (subtlexus.csv)
  • SUBTLEX$_{US}$ norms for US English
Joining Dataframes

- Sometimes different files/dataframes contain different variables relevant to the same observations
- Common scenario in mixed effects models context: Level-2 measurements are in a different file than Level-1 measurements

naming.csv: 1 row per trial
Each word appears in multiple rows

subtlexus.csv: Each word has only one row with its frequency
Joining Dataframes

- Sometimes different files/dataframes contain different variables relevant to the same observations
- Common scenario in mixed effects models context: Level-2 measurements are in a different file than Level-1 measurements

1 row per trial
Each subject has multiple rows

<table>
<thead>
<tr>
<th>SUBJECT</th>
<th>ITEM</th>
<th>CONDITION</th>
<th>CORRECT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Sentence1</td>
<td>Active</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>Sentence2</td>
<td>Passive</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>Sentence3</td>
<td>Active</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>Sentence4</td>
<td>Passive</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>Sentence1</td>
<td>Active</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>Sentence2</td>
<td>Passive</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>Sentence3</td>
<td>Active</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>Sentence4</td>
<td>Passive</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SUBJECT</th>
<th>READINGSPAN</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

Each subject has only one row with their Working Memory score
Joining Dataframes

- Sometimes different files/dataframes contain different variables relevant to the same observations.
- Common scenario in mixed effects models context: Level-2 variables are in a different file than Level-1 measurements.

allschools: 1 row per student

<table>
<thead>
<tr>
<th>School</th>
<th>Classroom</th>
<th>Student</th>
<th>HoursOfStudy</th>
<th>StudentSES</th>
<th>Pretest</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Jefferson</td>
<td>C001</td>
<td>S0001</td>
<td>1</td>
<td>0.7803573</td>
<td>0.5137800</td>
<td>1.5399431</td>
</tr>
<tr>
<td>2 Jefferson</td>
<td>C001</td>
<td>S0002</td>
<td>3</td>
<td>-0.2153623</td>
<td>0.2634907</td>
<td>1.3080398</td>
</tr>
<tr>
<td>3 Jefferson</td>
<td>C001</td>
<td>S0003</td>
<td>0</td>
<td>0.1290432</td>
<td>0.5232901</td>
<td>1.4550667</td>
</tr>
<tr>
<td>4 Jefferson</td>
<td>C001</td>
<td>S0004</td>
<td>3</td>
<td>1.6873593</td>
<td>0.3404230</td>
<td>0.6022264</td>
</tr>
<tr>
<td>5 Jefferson</td>
<td>C001</td>
<td>S0005</td>
<td>3</td>
<td>0.2196517</td>
<td>0.7866884</td>
<td>1.2517459</td>
</tr>
<tr>
<td>6 Jefferson</td>
<td>C001</td>
<td>S0006</td>
<td>5</td>
<td>-0.2931509</td>
<td>1.2862659</td>
<td>1.6046956</td>
</tr>
</tbody>
</table>

tutoruse.csv: Each class has only one row—did this class use the tutor or not?

<table>
<thead>
<tr>
<th>Classroom</th>
<th>Tutor</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>No</td>
</tr>
<tr>
<td>2</td>
<td>Yes</td>
</tr>
<tr>
<td>3</td>
<td>No</td>
</tr>
<tr>
<td>4</td>
<td>Yes</td>
</tr>
<tr>
<td>5</td>
<td>No</td>
</tr>
<tr>
<td>6</td>
<td>Yes</td>
</tr>
</tbody>
</table>
**Joining Dataframes**

- “Look up word frequency from the other dataframe”
- We can combine these dataframes if they have **at least one column in common**
- **Word** tells us which word was presented on an individual trial, and it also identifies the word in our database of word frequency

**naming.csv:**

<table>
<thead>
<tr>
<th>Subject</th>
<th>Word</th>
<th>YearsOfStudy</th>
<th>RT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>S1: memorandum</td>
<td>6</td>
<td>499.692</td>
</tr>
<tr>
<td>2</td>
<td>S1: shower</td>
<td>6</td>
<td>41.567</td>
</tr>
<tr>
<td>3</td>
<td>S1: suggest</td>
<td>6</td>
<td>144.895</td>
</tr>
<tr>
<td>4</td>
<td>S1: hair</td>
<td>6</td>
<td>119.509</td>
</tr>
<tr>
<td>5</td>
<td>S1: yell</td>
<td>6</td>
<td>251.718</td>
</tr>
<tr>
<td>6</td>
<td>S1: monorail</td>
<td>6</td>
<td>638.215</td>
</tr>
</tbody>
</table>

**subtlexus.csv:**

Each word appears in multiple rows.

<table>
<thead>
<tr>
<th>WordFreq</th>
<th>WordFreq</th>
<th>WordFreq</th>
</tr>
</thead>
<tbody>
<tr>
<td>the</td>
<td>6.1766</td>
<td>to</td>
</tr>
<tr>
<td></td>
<td>a</td>
<td>6.0175</td>
</tr>
<tr>
<td></td>
<td>you</td>
<td>6.3293</td>
</tr>
<tr>
<td></td>
<td>and</td>
<td>5.8343</td>
</tr>
<tr>
<td></td>
<td>it</td>
<td>5.9839</td>
</tr>
</tbody>
</table>
**Inner Join**

- `inner_join(naming, subtlexus, by='Word') -> naming2`
  - New dataframe (`naming2`) has both the columns from `naming` (Subject, YearsOfStudy, RT) and the columns from `subtlexus` (WordFreq)
  - Matches the observations using the **Word** column

**naming.csv:**
- 1 row per **trial**
- Each word appears in multiple rows

**subtlexus.csv:**
- Each word has only **one** row with its frequency
**Inner Join**

- `inner_join(naming, subtlexus, by='Word') -> naming2`
  - New dataframe (`naming2`) has both the columns from `naming` (`Subject, YearsOfStudy, RT`) and the columns from `subtlexus` (`WordFreq`)
  - Matches the observations using the `Word` column
  - Like VLOOKUP in Excel
Joins – Mismatching Column Names

- What if the columns have different names?
  - Not true in *this* dataset—just a hypothetical
  - **Item** in `naming` tells us which **Word** to look for in `subtlexus` ... how can we tell R that?
  - `inner_join(naming, subtlexus, by=c('Item'='Word'))` -> `naming2`

---

**naming**
```
Subject Item YearsOfStudy RT
1  S1  membrand  6 499.692
2  S1  shower    6 41.567
3  S1  suggest   6 144.895
4  S1  hair      6 119.509
5  S1  yell      6 251.718
6  S1  monorail  6 638.215
```

**subtlexus**
```
Word WordFreq
the 6.1766
to 6.0632
a 6.0175
you 6.3293
and 5.8343
it 5.9839
```
Other Types of Joins

- `nrow(naming)` 2940
- `nrow(naming2)` 2580
- Six words didn’t have a frequency measurement
- An **inner join** will drop rows that can’t be matched
- Alternative:
  - `naming2 <- left_join(naming, subtlexus, by='Word')`

Keep the rows in the first dataframe (`naming`) where we can’t find the matching WORD in the second dataframe (`subtlexus`)

Can you guess what happens to WordFreq for those trials?
Other Types of Joins

- \texttt{nrow(naming)} \quad 2940
- \texttt{nrow(naming2)} \quad 2580
- Six words didn’t have a frequency measurement
- An \texttt{inner join} will drop rows that can’t be matched
- Alternative:
  - \texttt{naming2} <- \texttt{left_join(naming, subtlexus, by='Word')}

Keep the rows in the first dataframe (\texttt{naming}) where we can’t find the matching WORD in the second dataframe (\texttt{subtlexus})

WordFreq will be NA (missing data) in these rows
Other Types of Joins

- `nrow(naming)`
- `nrow(naming2)`
- Six words didn’t have a frequency measurement
- An **inner join** will drop rows that can’t be matched
- A **left** or **right join** will keep every row in the first or second dataframe, respectively
- A **full join** keeps every row in *both* dataframes
  - `full_join(naming, subtlexus, by='Word')` -> `naming3`
  - Includes rows for every word in the English word frequency database, even ones not used in our experiment. We **DON’T** need or want that in this case.
Matching by Multiple Columns

- Sometimes, one column isn’t enough to uniquely match things across files/dataframes
- Can use multiple columns in join functions:
  - `inner_join(naming, subtlexus, by=c('Word', 'Country'))` -> naming2
  - This is a logical AND. Has to match both Word and Country

Imagine doing our task in both the US and UK. Word frequency differs somewhat between American English & British English, so now we need both Word and Country to look up the frequency.
Crossed Random Effects

• Now that we’ve added the word frequency to our dataframe, let’s add it to the model
  • `model2 <- lmer(RT ~ 1 + YearsOfStudy * WordFreq + (1|Subject) + (1|Word),
  data=naming2)`

Greater Subject variance than Item variance. Typical in experiments because we often design items to be similar.
Week 6.1: Crossed Random Effects

- Crossed Random Effects Analysis
  - Example Design
  - Joining Dataframes

- Random Slopes in Crossed Designs
  - Between-Subjects & Within-Subjects Designs
    - Between-Items & Within-Items Designs
    - When are Random Slopes Necessary?
    - Practice Activity
What kind of random slopes might be relevant here?

To answer this question, we need to understand how “between-subjects variables” differ from “within-subjects variables”
**Between vs. Within Subjects**

- In an experimental design, some variables are:
  - **Between-Subjects variables**: Each subject is in 1 and only 1 group ... or has 1 and only 1 value
    - Randomly assigned to drug 1 vs. drug 2 vs. placebo
    - Demographic variables; e.g., SES
    - Cognitive/linguistic differences (e.g., working mem. score)
    - “Between subjects” because differences in this variable are only seen between one subject and another

---

**SUBJECT 10’s DATA**
- Native speaker
  - Trial 1: Correct
  - Trial 2: Correct
  - Trial 3: Incorrect

**SUBJECT 11’s DATA**
- Non-native speaker
  - Trial 1: Incorrect
  - Trial 2: Correct
  - Trial 3: Incorrect
Between vs. Within Subjects

• In an experimental design, some variables are:
  ➢ **Between-Subjects variables**: Each subject is in 1 and only 1 group ... or has 1 and only 1 value
    • Randomly assigned to drug 1 vs. drug 2 vs. placebo
    • Demographic variables; e.g., SES
    • Cognitive/linguistic differences (e.g., working mem. score)
    • “Between subjects” because differences in this variable are only seen between one subject and another
  ➢ **Within-Subjects variables**: Same subject sees more than 1 condition or has >1 value
    • Same subject sees both congruent (green) and incongruent (blue) Stroop trials
    • Values that vary w/in a study, e.g., # of previous trials
    • Variables where you’d use a repeated measures ANOVA
    • “Within-subjects” because you can see differences in this variable even within a single subject
**Between vs. Within Subjects**

- In an experimental design, some variables are:
  - **Between-Subjects variables**: Each subject is in 1 and only 1 group … or has 1 and only 1 value
    - Randomly assigned to
    - Demographic variables; e.g., SES
    - Cognitive/linguistic differences (e.g., working mem. score)
  - “Between subjects” because differences in this variable are only seen between one subject and another

- **Within-Subjects variables**: Same subject sees more than 1 condition or has >1 value
  - Same subject sees both congruent (green) and incongruent (blue) Stroop trials
  - Values that vary w/in a study, e.g., # of previous trials
  - Variables where you’d use a repeated measures ANOVA
  - “Within-subjects” because you can see differences in this variable even within a single subject

---

**SUBJECT 12’S DATA**

- Trial 1: Congruent Stroop, 655 ms
- Trial 2: Incongruent Stroop, 512 ms
- Trial 3: Incongruent Stroop, 711 ms
- Trial 4: Congruent Stroop: 642 ms
The same variable could end up as between- or within-subjects, depending on experimental design.

I’m interested in maintenance rehearsal (repetition) vs. elaborative rehearsal (relating to other concepts).

Half of my participants study second-language vocab using maintenance rehearsal, and half study words using elaborative rehearsal. 

Each participant studies some words with maintenance rehearsal and some with elaborative rehearsal.
Between vs. Within Subjects

• How about in our naming dataset?

• **Years of study** is…
  • This experiment takes place at a single point in time, so the number of years a subject has been studying English is fixed and never varies
  • Between subjects

• **Word frequency** is…
  • Each subject sees high-, medium- and low-frequency words
  • Within subjects
When are Random Slopes Appropriate?

• If the random effect is subject...

• Random slopes for subjects appropriate for:
  • Within-subjects variables
  • We can draw a regression line for each subject
  • Calculate the effect “within” each subject

<table>
<thead>
<tr>
<th>Subject 1 High Frequency</th>
<th>Subject 1 Low Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject 2 High Frequency</td>
<td>Subject 2 Low Frequency</td>
</tr>
</tbody>
</table>

• Random slopes for subjects inappropriate for:
  • Between-subjects variables
  • Can’t draw a regression line within each subject

| Subject 1: 6 years of study | Subject 2: 2 years of study |
Random Slopes: Implementation

• Remember, previous model was:
  • `model2 <- lmer(RT ~ 1 + YearsOfStudy * WordFreq + (1|Subject) + (1|Word), data=naming2)`

• What variable varies within subjects? Try adding a random slope for it
  • `model3 <- lmer(RT ~ 1 + YearsOfStudy * WordFreq + ???????????????????? + (1|Word), data=naming2)`
Random Slopes: Implementation

- Remember, previous model was:
  - `model2 <- lmer(RT ~ 1 + YearsOfStudy * WordFreq + (1|Subject) + (1|Word), data=naming2)`

- What variable varies within subjects? Try adding a random slope for it
  - `model3 <- lmer(RT ~ 1 + YearsOfStudy * WordFreq + (1+WordFreq|Subject) + (1|Word), data=naming2)`

Subjects differ in their intercept (baseline RT)
Subjects differ in the effect of word frequency on their RTs

Again, miniature model formula for things we think will vary by subjects
Random Slopes: Implementation

• Remember, previous model was:
  • model2 <- lmer(RT ~ 1 + YearsOfStudy * WordFreq + (1|Subject) + (1|Word), data=naming2)

• What variable varies within subjects? Try adding a random slope for it
  • model3 <- lmer(RT ~ 1 + YearsOfStudy * WordFreq + (1+WordFreq|Subject) + (1|Word), data=naming2)

Random effects:

<table>
<thead>
<tr>
<th>Groups</th>
<th>Name</th>
<th>Variance</th>
<th>Std.Dev.</th>
<th>Corr</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject</td>
<td>(Intercept)</td>
<td>1097.0</td>
<td>33.12</td>
<td></td>
</tr>
<tr>
<td>Word</td>
<td>(Intercept)</td>
<td>264.2</td>
<td>16.25</td>
<td></td>
</tr>
<tr>
<td>Residual</td>
<td></td>
<td>5431.9</td>
<td>73.70</td>
<td></td>
</tr>
</tbody>
</table>

Number of obs: 2580, groups: Subject, 60; Word, 43

Fixed effects:

|                     | Estimate | Std. Error | df | t value | Pr(>|t|) |
|---------------------|----------|------------|----|---------|---------|
| (Intercept)         | 936.208  | 19.138     | 79.484 | 48.918  | < 2e-16 *** |
| YearsOfStudy        | -41.663  | 4.258      | 57.999 | -9.786  | 6.89e-14 *** |
| WordFreq            | -123.183 | 6.804      | 76.274 | -18.105 | < 2e-16 *** |
| YearsOfStudy:WordFreq| 1.542    | 1.564      | 58.003 | 0.986   | 0.328    |
**Random Slopes**

- Original model says that subjects vary in baseline RT
  - Random intercept
- And that a 1-unit change in word frequency $\approx 120$ ms decrease in RT
  - Fixed effect across subjects
Random Slopes

- Original model says that subjects vary in baseline RT
  - Random intercept
- And that a 1-unit change in word frequency ≈ 120 ms decrease in RT
  - Fixed effect across subjects
- Differences in slope capture how individual people vary in sensitivity to word frequency
  - Random slope
- Such differences may correlate with baseline
Week 6.1: Crossed Random Effects

- Crossed Random Effects Analysis
  - Example Design
  - Joining Dataframes

- Random Slopes in Crossed Designs
  - Between-Subjects & Within-Subjects Designs
  - Between-Items & Within-Items Designs
    - When are Random Slopes Necessary?
    - Practice Activity
Between vs. Within Items

• We can draw a similar distinction between
  ➢ **Between-Items variables**: Each item appears in 1 and only 1 condition … or has 1 and only 1 value
    • Visual complexity of pictures
    • One set of sentences is used in our “plausible” condition and a completely different set is used in our “implausible” condition
    • “Between items” because differences in this variable are only seen between one item and another
  ➢ **Within-Items variables**: Same item appears in more than 1 condition or has >1 value
    • Same science facts presented (across subjects) in either an elaborative- or maintenance-rehearsal condition
    • We manipulate the same fictitious resume to have either a stereotypically African-American or European-American name
Between vs. Within Items

• Try updating the most recent model with random slope(s) for the within-items variable(s)
  • Hint: Think about whether each variable is constant for a particular Word or not

• The final full model:
  • model.Full <- lmer(RT ~ 1 + YearsOfStudy * WordFreq + (1 + WordFreq|Subject) + ??????????????????? , data=naming2)
Between vs. Within Items

• Try updating the most recent model with random slope(s) for the within-items variable(s)
  • Hint: Think about whether each variable is constant for a particular Word or not

• The final full model:
  • `model.Full <- lmer(RT ~ 1 + YearsOfStudy * WordFreq + (1 + WordFreq|Subject) + (1 + YearsOfStudy|Word), data=naming2)`
Between vs. Within Items

- Why didn’t we include a random slope of \textit{WordFreq} by items?
- Each word has a \textit{constant} frequency (within this experiment)
  - Doesn’t make sense to discuss effects of word frequency \textit{within} an item
  - No random slope of frequency by items
- But, each item IS presented to different subjects who differ in their YearsOfStudy
Week 6.1: Crossed Random Effects

- Crossed Random Effects Analysis
  - Example Design
  - Joining Dataframes

- Random Slopes in Crossed Designs
  - Between-Subjects & Within-Subjects Designs
  - Between-Items & Within-Items Designs

- When are Random Slopes Necessary?
  - Practice Activity
When Are Random Slopes Necessary?

- Remember that failing to account for clustering could inflate our Type I error rate

- With factorial experiments, standard is to at least try to include all random slopes we can (Barr et al., 2013)
  - **Maximal** random effects structure
  - Expectation is that we assume that subjects could vary in, say, their WordFreq effect—that’s why we ran more than one subject
  - If this does not converge, simplify using strategies we discussed last class (e.g., removing correlation parameters)
When Are Random Slopes Necessary?

• Remember that failing to account for clustering could inflate our Type I error rate

• In more observational data, not necessarily the case you’d include all possible random slopes
  • What’s theoretically relevant or expected?
The View Ahead

• We’ve covered the basics of fitting a mixed effects model
  • You will practice this in Wednesday’s lab

• But, we have only used continuous variables

• Next 3 weeks: **Categorical** variables (factors)
  • Next two weeks: Categorical **predictors** (IVs)
    • e.g., experimental vs. control condition; race/ethnicity
  • After that: Categorical **outcomes** (DVs)
    • e.g., recalled vs. didn’t recall a science fact; did or didn’t graduate high school
Week 6.1: Crossed Random Effects

- Crossed Random Effects Analysis
  - Example Design
  - Joining Dataframes

- Random Slopes in Crossed Designs
  - Between-Subjects & Within-Subjects Designs
  - Between-Items & Within-Items Designs
  - When are Random Slopes Necessary?

- Practice Activity