**Course Business**

- LOTS of data on CourseWeb for this week
  - Cognitive Tutor use in schools
  - Word processing (“lexical decision”) task

- Install package **tidyverse**

- Last day of lecture!

- Course evaluation (OMET) survey available
  - E-mailed to you and also on CourseWeb
Course Business

• Final project presentation schedule:
  • Dec 5\textsuperscript{th}: Doug, Jenny, Kelly, Kole, Rob, Zac
  • Dec 12\textsuperscript{th}: Ciara, Griff, Kori, Lauren, Lin, Rebecca
  • Details on CourseWeb

• Guidelines:
  • Presenters: About 10 min. presentation + 5 min. questions & answers
  • Audience: Plan to ask at least 1 question over the two days

• Can use my laptop or your own
  • If you e-mail me file beforehand, I can check if it displays correctly on my computer
Week 13: Data Management & Level-2 Variables

- Finish Power
  - Your Own Power Analysis
  - Influences on Power
- Data Management in R
  - Combining More of the Same
  - Joins
  - Reshaping Data
- Level-2 Fixed & Random Effects
  - Understanding Level-2 Variables
    - Level-2 Random Effects
    - BLUPs
    - Level-2 Fixed Effects
  - Continuous or categorical?
    - Median splits
    - Extreme groups design
  - Good measurement
    - Reliability
    - Validity
A positive psychology lab is examining how feelings of subjective well being (SWB) vary over the course of the typical workweek. As a pilot, 20 participants come to the lab on Monday to participate in the first session and to get an app for their phones. We then use the app to poll the participants (i.e., ecological momentary assessment; EMA) on their SWB (rated 1 to 7) once each of the remaining days of the week.

We run the following model:

```r
model1 <- lmer(SWB ~ 1 + DayOfWeek + SessionNumber + (1 + DayOfWeek + SessionNumber | Subject), data=positivepsych)
```

However, we receive the following error message:

```
fixed-effect model matrix is rank deficient so dropping 1 column / coefficient
```
Distributed Practice: The Final Chapter

• The lab disagrees on what we should do:
  • Andre says, “Let’s increase the maxfun parameter to allow the model more chances to converge.”
  • Bill says, “DayOfWeek and SessionNumber are perfectly confounded; we can fix this error by removing one of them.”
  • Caitlin says, “The random-effects structure is probably too complex. Let’s simplify it by removing the correlation parameters by using ||Subject instead of |Subject”
  • Donghee says, “We can deal with the low sample size by computing the empirical logit and using that as our new DV.”
Distributed Practice: The Final Chapter

• The lab disagrees on what we should do:
  • Andre says, “Let’s increase the maxfun parameter to allow the model more chances to converge.”
    • This isn’t a failure to converge.
    • And, simply adding more iterations often does not fix convergence errors.
  Bill says, “DayOfWeek and SessionNumber are perfectly confounded; we can fix this error by removing one of them.”

• Caitlin says, “The random-effects structure is probably too complex. Let’s simplify it by removing the correlation parameters by using ||Subject instead of |Subject”
  • This would indeed simplify the random effects structure, but there is no reason to think it’s a problem—it’s not what the error message is about
• Donghee says, “We can deal with the low sample size by computing the empirical logit and using that as our new DV.”
  • The empirical logit is only relevant for a binomial DV
Dr. Hourihan is studying the cross-race effect in memory—the tendency for people to have better memory for faces of their own race. She recruits 30 African-American and 30 European-American participants. Each studies a series of 40 faces: 20 of their own race, 20 of the other race. Then, they are presented with a recognition memory test in which they see the 40 studied faces & 40 new faces and have to judge whether each face was seen before or not. The table below shows the % of times Dr. Hourihan’s subjects judged the faces as seen:

<table>
<thead>
<tr>
<th>Studied Face</th>
<th>Same-Race</th>
<th>Different-Race</th>
</tr>
</thead>
<tbody>
<tr>
<td>Studied Face</td>
<td>70%</td>
<td>60%</td>
</tr>
<tr>
<td>New Face</td>
<td>30%</td>
<td>40%</td>
</tr>
</tbody>
</table>

- Do same- vs. other-race faces differ in sensitivity?
- Do same- vs. other-race faces differ in response bias?
Dr. Hourihan is studying the cross-race effect in memory—the tendency for people to have better memory for faces of their own race. She recruits 30 African-American and 30 European-American participants. Each studies a series of 40 faces: 20 of their own race, 20 of the other race. Then, they are presented with a recognition memory test in which they see the 40 studied faces & 40 new faces and have to judge whether each face was seen before or not. The table below shows the % of times Dr. Hourihan’s subjects judged the faces as seen:

<table>
<thead>
<tr>
<th></th>
<th>Same-Race</th>
<th>Different-Race</th>
</tr>
</thead>
<tbody>
<tr>
<td>Studied Face</td>
<td>70%</td>
<td>60%</td>
</tr>
<tr>
<td>New Face</td>
<td>30%</td>
<td>40%</td>
</tr>
</tbody>
</table>

Do same- vs. other-race faces differ in sensitivity?
- Yes. Bigger responding difference between studied & new faces for same-race faces.
Dr. Hourihan is studying the cross-race effect in memory—the tendency for people to have better memory for faces of their own race. She recruits 30 African-American and 30 European-American participants. Each studies study a series of 40 faces: 20 of their own race, 20 of the other race. Then, they are presented with a recognition memory test in which they see the 40 studied faces & 40 new faces and have to judge whether each face was seen before or not. The table below shows the % of times Dr. Hourihan’s subjects judged the faces as seen:

<table>
<thead>
<tr>
<th></th>
<th>Same-Race</th>
<th>Different-Race</th>
</tr>
</thead>
<tbody>
<tr>
<td>Studied Face</td>
<td>70%</td>
<td>60%</td>
</tr>
<tr>
<td>New Face</td>
<td>30%</td>
<td>40%</td>
</tr>
<tr>
<td></td>
<td>Average = 50%</td>
<td>Average = 50%</td>
</tr>
</tbody>
</table>

Do same- vs. other-race faces differ in response bias?

No. Same overall rate of “yes” responses.
Week 13: Data Management & Level-2 Variables

- Finish Power
  - Your Own Power Analysis
  - Influences on Power
- Data Management in R
  - Combining More of the Same
  - Joins
  - Reshaping Data
- Level-2 Fixed & Random Effects
  - Understanding Level-2 Variables
    - Level-2 Random Effects
    - BLUPs
    - Level-2 Fixed Effects
  - Continuous or categorical?
    - Median splits
    - Extreme groups design
  - Good measurement
    - Reliability
    - Validity
Your Own Power Analysis

• Rationale behind power analyses:
  • Can we detect the kind & size of effect we’re interested in?
  • What sample size would we need?

• In practice:
  • We can’t control effect size; it’s a property of nature
  • \( \alpha \) is usually fixed (e.g., at .05) by convention
  • But, we can control our sample size \( n \)!

• So:
  • Determine desired power (often .80)
  • Estimate the effect size(s)
  • Calculate the necessary sample size \( n \)
• Power for ANOVAs can be easily found from tables
  • Simpler design. Only 1 random effect (at most)
• More complicated for mixed effect models
Monte Carlo Methods

• Remember the definition of power?
  • The probability of observing a significant effect in our sample if the effect *truly exists in the population*
  • What if we knew for a fact that the effect existed in a particular population?
  • Then, a measure of *power* is how often we get a significant result in a sample (of our intended \( n \))
  • Observe a significant effect 10 samples out of 20 = 50% of the time = power of .50
  • Observe a significant effect 300 samples out of 1000 = 30% of the time = power of .30
  • Observe a significant effect 800 samples out of 1000 = 80% of the time = power of .80
Monte Carlo Methods

• Remember the definition of power?
  • The probability of observing a significant effect in our sample if the effect truly exists in the population
  • What if we knew for a fact that the effect existed in a particular population?
  • Then, a measure of power is how often we get a significant result in a sample (of our intended $n$)

Great, but where am I ever going to find data where I know exactly what the population parameters are?
Monte Carlo Methods

• Remember the definition of power?
  • The probability of observing a significant effect in our sample if the effect truly exists in the population
  • What if we knew for a fact that the effect existed in a particular population?
  • Then, a measure of power is how often we get a significant result in a sample (of our intended \( n \))

• Solution: We create (“simulate”) the data.
**Data Simulation**

- Set some plausible population parameters *(effect size, subject variance, item var., etc.)*

  Set population parameters
  - Mean = 723 ms
  - Group difference = 100 ms
  - Subject var = 30

- Since we are *creating* the data…
  - We can choose the population parameters
  - We know we exactly what they are
Data Simulation

- Create (“simulate”) a random sample drawn from this population

Set population parameters
Mean = 723 ms
Group difference = 100 ms
Subject var = 30

Create a random sample from these data
N subjects = 20
N items = 40

- Like most samples, the sample statistics will not exactly match the population parameters
  - It’s randomly generated
- But, the difference is we know what the population is like & that there IS an effect
**Data Simulation**

- Now, fit our planned mixed-effects model to this **sample** of simulated data to get **one** result

  - **Set population parameters**
    - Mean = 723 ms
    - Group difference = 100 ms
    - Subject var = 30

  - **Create a random sample** from these data
    - \(N\) subjects = 20
    - \(N\) items = 40

  - **Run our planned model and see if we get a significant result**

- **Might get a significant result**
  - Correctly detected the effect in the **population**

- **Might get a non-significant result**
  - Type II error – missed an effect that **really exists in the population**
Monte Carlo Methods

- If we do this repeatedly, we will get *multiple* significance tests, each on a different sample

Set population parameters
Mean = 723 ms
Group difference = 100 ms
Subject var = 30

Create a random sample from these data
$N$ subjects = 20
$N$ items = 40

Run our planned model and see if we get a significant result

Repeat with a new sample from the same population

Outcomes:
- Sample 1: $p < .05$ (Yes)
- Sample 2: $p = .23$ (No)
- Sample 3: $p < .05$ (Yes)
- Sample 4: $p = .14$ (No)

Detected the effect $\frac{1}{2}$ of the time: Power = .50
Monte Carlo Methods

- If we do this repeatedly, we will get *multiple* significance tests, each on a different sample

  - Set population parameters
    - Mean = 723 ms
    - Group difference = 100 ms
    - Subject var = 30

  - Create a random sample from these data
    - N subjects = 20
    - N items = 40

  - Run our planned model and see if we get a significant result

  - Hmm, that power wasn’t very good 😞

  - Repeat with a new sample from the same population
Monte Carlo Methods

- If we do this repeatedly, we will get *multiple* significance tests, each on a different sample

  - Set population parameters
    - Mean = 723 ms
    - Group difference = 100 ms
    - Subject var = 30

  - Create a random sample from these data
    - N subjects = 60
    - N items = 40

  - Run our planned model and see if we get a significant result

  - Repeat with a new sample from the same population

- Hmm, that power wasn’t very good 😞

- Let’s increase the number of subjects and run a new simulation to see what our power is like now
Monte Carlo Methods

- If we do this repeatedly, we will get *multiple* significance tests, each on a different sample

  - Set population parameters
    - Mean = 723 ms
    - Group difference = 100 ms
    - Subject var = 30

  - Create a random sample from these data
    - N subjects = 60
    - N items = 40

  - Run our planned model and see if we get a significant result

- Goal: Find the sample size(s) that let you detect the effect at least 80% of the time (or whatever your desired *power* is)
  - Will 40 subjects in each of 5 schools suffice?
  - What about 50 subjects in each of 10 schools?
Week 13: Data Management & Level-2 Variables

- Finish Power
  - Your Own Power Analysis
  - Influences on Power
- Data Management in R
  - Combining More of the Same
  - Joins
  - Reshaping Data
- Level-2 Fixed & Random Effects
  - Understanding Level-2 Variables
    - Level-2 Random Effects
    - BLUPs
    - Level-2 Fixed Effects
  - Continuous or categorical?
    - Median splits
    - Extreme groups design
  - Good measurement
    - Reliability
    - Validity
Influences on Power

• So what makes for a powerful design?
• Things that *increase* power:
  • Larger **effect size estimates** for the fixed effects
    • Bigger things are easier to find
  • Larger **sample size** (at any level)
    • More data = more confidence
    • This one we can control
    • Increasing sample size at a higher level (e.g., subjects rather than time points within subjects) is more effective
• Variance of **independent variables**
  • Easier to see an effect of income on happiness if people *vary* in their income
  • *Hard* to test effect of “number of fingers on your hand”
  • With a categorical variable, would prefer to have an equal # of observations in each condition—most information
Influences on Power

• So what makes for a powerful design?
• Things that decrease power:
  • Larger variance of random effects
    • More differences between people (noise) make it harder to see what’s consistent
  • Larger error variance
    • Again, more noise = harder to see consistent effects
    • May be able to reduce if you can add covariates / control variables
Week 13: Data Management & Level-2 Variables

- Finish Power
  - Your Own Power Analysis
  - Influences on Power

- Data Management in R
  - Combining More of the Same
  - Joins
  - Reshaping Data

- Level-2 Fixed & Random Effects
  - Understanding Level-2 Variables
    - Level-2 Random Effects
    - BLUPs
    - Level-2 Fixed Effects
  - Continuous or categorical?
    - Median splits
    - Extreme groups design
  - Good measurement
    - Reliability
    - Validity
Week 13: Data Management & Level-2 Variables

- Lots of data today on CourseWeb today (we’ll be talking about how to combine it):
  - school1.csv
  - school2.csv
  - school3.csv
    - Performance on math SAT in three different schools
  - tutoruse.csv
    - Whether each classroom used a computer adaptive math tutor or not.
      Stored in a separate file so the experimenter is blind to this
  - lexicaldecision.csv
    - Cognitive task measuring word processing. See a string of letters,
      decide if it's
  - subtlexus.csv
Look at the **school1**, **school2**, **school3** dataframes

- How are they similar? How are they different?
- Pre-test & post-test math SAT for each 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>Jefferson:300</td>
<td>C001 : 30</td>
<td>S0001 : 1</td>
<td>Min. : 0.000</td>
<td>Min. : -3.21423</td>
<td>Min. : 268.0</td>
<td>Min. : 352.0</td>
</tr>
<tr>
<td></td>
<td>C002 : 30</td>
<td>S0002 : 1</td>
<td>1st Qu. : 1.000</td>
<td>1st Qu. : -0.69725</td>
<td>1st Qu. : 458.8</td>
<td>1st Qu. : 540.8</td>
</tr>
<tr>
<td></td>
<td>C003 : 30</td>
<td>S0003 : 1</td>
<td>Median : 3.000</td>
<td>Median : 0.09570</td>
<td>Median : 538.5</td>
<td>Median : 615.5</td>
</tr>
<tr>
<td></td>
<td>C004 : 30</td>
<td>S0004 : 1</td>
<td>Mean : 2.613</td>
<td>Mean : 0.07882</td>
<td>Mean : 534.0</td>
<td>Mean : 611.9</td>
</tr>
<tr>
<td></td>
<td>C005 : 30</td>
<td>S0005 : 1</td>
<td>3rd Qu. : 4.000</td>
<td>3rd Qu. : 0.51511</td>
<td>3rd Qu. : 605.2</td>
<td>3rd Qu. : 686.5</td>
</tr>
<tr>
<td></td>
<td>C006 : 30</td>
<td>S0006 : 1</td>
<td>Max. : 5.000</td>
<td>Max. : 2.37009</td>
<td>Max. : 800.0</td>
<td>Max. : 800.0</td>
</tr>
<tr>
<td>(Other):120</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<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>Hoover:300</td>
<td>C011 : 30</td>
<td>S0301 : 1</td>
<td>Min. : -2.72910</td>
<td>Min. : 0.000</td>
<td>Min. : 200.0</td>
<td>Min. : 268.0</td>
</tr>
<tr>
<td></td>
<td>C012 : 30</td>
<td>S0302 : 1</td>
<td>1st Qu. : -0.62818</td>
<td>1st Qu. : 1.000</td>
<td>1st Qu. : 307.0</td>
<td>1st Qu. : 407.8</td>
</tr>
<tr>
<td></td>
<td>C013 : 30</td>
<td>S0303 : 1</td>
<td>Median : 0.03683</td>
<td>Median : 3.000</td>
<td>Median : 383.5</td>
<td>Median : 478.5</td>
</tr>
<tr>
<td></td>
<td>C014 : 30</td>
<td>S0304 : 1</td>
<td>Mean : 0.01090</td>
<td>Mean : 2.557</td>
<td>Mean : 380.3</td>
<td>Mean : 482.0</td>
</tr>
<tr>
<td></td>
<td>C015 : 30</td>
<td>S0305 : 1</td>
<td>3rd Qu. : 0.71810</td>
<td>3rd Qu. : 4.000</td>
<td>3rd Qu. : 446.2</td>
<td>3rd Qu. : 551.5</td>
</tr>
<tr>
<td></td>
<td>C016 : 30</td>
<td>S0306 : 1</td>
<td>Max. : 2.76739</td>
<td>Max. : 5.000</td>
<td>Max. : 700.0</td>
<td>Max. : 800.0</td>
</tr>
<tr>
<td>(Other):120</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>School</th>
<th>Classroom</th>
<th>Student</th>
<th>HoursOfStudy</th>
<th>StudentSES</th>
<th>VerbalSAT</th>
<th>Pretest</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Harding:300</td>
<td>C021 : 30</td>
<td>S0601 : 1</td>
<td>Min. : 0.000</td>
<td>Min. : -3.19184</td>
<td>Min. : 200.0</td>
<td>Min. : 200.0</td>
<td>Min. : 270.0</td>
</tr>
<tr>
<td></td>
<td>C022 : 30</td>
<td>S0602 : 1</td>
<td>1st Qu. : 1.000</td>
<td>1st Qu. : -0.60990</td>
<td>1st Qu. : 307.8</td>
<td>1st Qu. : 356.8</td>
<td>1st Qu. : 441.8</td>
</tr>
<tr>
<td></td>
<td>C023 : 30</td>
<td>S0603 : 1</td>
<td>Median : 2.000</td>
<td>Median : 0.05070</td>
<td>Median : 402.0</td>
<td>Median : 435.0</td>
<td>Median : 521.5</td>
</tr>
<tr>
<td></td>
<td>C024 : 30</td>
<td>S0604 : 1</td>
<td>Mean : 2.39</td>
<td>Mean : 0.06499</td>
<td>Mean : 415.3</td>
<td>Mean : 421.8</td>
<td>Mean : 509.9</td>
</tr>
<tr>
<td></td>
<td>C025 : 30</td>
<td>S0605 : 1</td>
<td>3rd Qu. : 14.00</td>
<td>3rd Qu. : 0.78731</td>
<td>3rd Qu. : 513.8</td>
<td>3rd Qu. : 492.0</td>
<td>3rd Qu. : 576.2</td>
</tr>
<tr>
<td></td>
<td>C026 : 30</td>
<td>S0606 : 1</td>
<td>Max. : 5.00</td>
<td>Max. : 2.45839</td>
<td>Max. : 800.0</td>
<td>Max. : 770.0</td>
<td>Max. : 800.0</td>
</tr>
<tr>
<td>(Other):120</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Look at the `school1`, `school2`, `school3` dataframes.

- How are they similar? How are they different?
- Pre-test & post-test math SAT for each student

### Columns not always in same order

<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>Jefferson:300</td>
<td>C001 : 30</td>
<td>S0001 : 1</td>
<td>1</td>
<td>Min. : 0.000, Min. : -3.21423</td>
<td>1st Qu.: 1.000, 1st Qu.: -0.69725</td>
<td>Median : 3.000, Median : 0.09570</td>
</tr>
<tr>
<td></td>
<td>C002 : 30</td>
<td>S0002 : 1</td>
<td>1</td>
<td>Mean : 2.613, Mean : 0.07882</td>
<td>3rd Qu.: 4.000, 3rd Qu.: 0.51511</td>
<td>Max. : 5.000, Max. : 2.37009</td>
</tr>
<tr>
<td></td>
<td>C003 : 30</td>
<td>S0003 : 1</td>
<td>1</td>
<td>Min. : 2.72910, Min. : 0.000</td>
<td>1st Qu.: 0.62818, 1st Qu.: 1.000</td>
<td>Median : 0.03683, Median : 3.000</td>
</tr>
<tr>
<td></td>
<td>C004 : 30</td>
<td>S0004 : 1</td>
<td>1</td>
<td>Mean : 0.01909, Mean : 2.557</td>
<td>3rd Qu.: 0.71810, 3rd Qu.: 4.000</td>
<td>Max. : 2.76739, Max. : 5.000</td>
</tr>
<tr>
<td></td>
<td>C005 : 30</td>
<td>S0005 : 1</td>
<td>1</td>
<td>Min. : 0.000, Min. : -3.19184</td>
<td>1st Qu.: 1.000, 1st Qu.: -0.60990</td>
<td>Median : 2.000, Median : 0.05070</td>
</tr>
<tr>
<td></td>
<td>C006 : 30</td>
<td>S0006 : 1</td>
<td>1</td>
<td>Mean : 2.39, Mean : 0.06499</td>
<td>3rd Qu.: 4.000, 3rd Qu.: 0.78731</td>
<td>Max. : 5.000, Max. : 2.45839</td>
</tr>
<tr>
<td>(Other):120</td>
<td>(Other):294</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>School</th>
<th>Classroom</th>
<th>Student</th>
<th>HoursOfStudy</th>
<th>StudentSES</th>
<th>VerbalSAT</th>
<th>Pretest</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hoover:300</td>
<td>C011 : 30</td>
<td>S0301 : 1</td>
<td>1</td>
<td>Min. : 2.000, Min. : 0.000</td>
<td>1st Qu.: 0.62818, 1st Qu.: 1.000</td>
<td>Median : 0.03683, Median : 3.000</td>
<td></td>
</tr>
<tr>
<td></td>
<td>C012 : 30</td>
<td>S0302 : 1</td>
<td>1</td>
<td>Mean : 0.01909, Mean : 2.557</td>
<td>3rd Qu.: 0.71810, 3rd Qu.: 4.000</td>
<td>Max. : 2.76739, Max. : 5.000</td>
<td></td>
</tr>
<tr>
<td></td>
<td>C013 : 30</td>
<td>S0303 : 1</td>
<td>1</td>
<td>Min. : 2.72910, Min. : 0.000</td>
<td>1st Qu.: 0.62818, 1st Qu.: 1.000</td>
<td>Median : 0.03683, Median : 3.000</td>
<td></td>
</tr>
<tr>
<td></td>
<td>C014 : 30</td>
<td>S0304 : 1</td>
<td>1</td>
<td>Mean : 0.01909, Mean : 2.557</td>
<td>3rd Qu.: 0.71810, 3rd Qu.: 4.000</td>
<td>Max. : 2.76739, Max. : 5.000</td>
<td></td>
</tr>
<tr>
<td></td>
<td>C015 : 30</td>
<td>S0305 : 1</td>
<td>1</td>
<td>Min. : 2.72910, Min. : 0.000</td>
<td>1st Qu.: 0.62818, 1st Qu.: 1.000</td>
<td>Median : 0.03683, Median : 3.000</td>
<td></td>
</tr>
<tr>
<td></td>
<td>C016 : 30</td>
<td>S0306 : 1</td>
<td>1</td>
<td>Mean : 0.01909, Mean : 2.557</td>
<td>3rd Qu.: 0.71810, 3rd Qu.: 4.000</td>
<td>Max. : 2.76739, Max. : 5.000</td>
<td></td>
</tr>
<tr>
<td>(Other):120</td>
<td>(Other):294</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>School</th>
<th>Classroom</th>
<th>Student</th>
<th>HoursOfStudy</th>
<th>StudentSES</th>
<th>VerbalSAT</th>
<th>Pretest</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Harding:300</td>
<td>C021 : 30</td>
<td>S0601 : 1</td>
<td>1</td>
<td>Min. : 0.000, Min. : -3.19184</td>
<td>1st Qu.: 1.000, 1st Qu.: -0.60990</td>
<td>Median : 2.000, Median : 0.05070</td>
<td></td>
</tr>
<tr>
<td></td>
<td>C022 : 30</td>
<td>S0602 : 1</td>
<td>1</td>
<td>Mean : 2.39, Mean : 0.06499</td>
<td>3rd Qu.: 4.000, 3rd Qu.: 0.78731</td>
<td>Max. : 5.000, Max. : 2.45839</td>
<td></td>
</tr>
<tr>
<td></td>
<td>C023 : 30</td>
<td>S0603 : 1</td>
<td>1</td>
<td>Min. : 0.000, Min. : -3.19184</td>
<td>1st Qu.: 1.000, 1st Qu.: -0.60990</td>
<td>Median : 2.000, Median : 0.05070</td>
<td></td>
</tr>
<tr>
<td></td>
<td>C024 : 30</td>
<td>S0604 : 1</td>
<td>1</td>
<td>Mean : 2.39, Mean : 0.06499</td>
<td>3rd Qu.: 4.000, 3rd Qu.: 0.78731</td>
<td>Max. : 5.000, Max. : 2.45839</td>
<td></td>
</tr>
<tr>
<td></td>
<td>C025 : 30</td>
<td>S0605 : 1</td>
<td>1</td>
<td>Min. : 0.000, Min. : -3.19184</td>
<td>1st Qu.: 1.000, 1st Qu.: -0.60990</td>
<td>Median : 2.000, Median : 0.05070</td>
<td></td>
</tr>
<tr>
<td></td>
<td>C026 : 30</td>
<td>S0606 : 1</td>
<td>1</td>
<td>Mean : 2.39, Mean : 0.06499</td>
<td>3rd Qu.: 4.000, 3rd Qu.: 0.78731</td>
<td>Max. : 5.000, Max. : 2.45839</td>
<td></td>
</tr>
<tr>
<td>(Other):120</td>
<td>(Other):294</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
```
**school1, school2, school3**

- Look at the **school1, school2, school3** dataframes
  - How are they similar? How are they different?
  - Pre-test & post-test math SAT for each student

```r
\[
\text{Only school3 has verbal SAT scores}
\]
**bind_rows()**

- Overall, this is similar information, so let’s combine it all
- Paste together the rows from two (or more) dataframes to create a new one:
  - `allschools <- bind_rows(school1, school2, school3)`

- Useful when observations are spread across files
  - Or, to create a dataframe that consists of 2 subsets
  - “More of the same”
**bind_rows(): Results**

- Resulting dataframe:

<table>
<thead>
<tr>
<th>School</th>
<th>Classroom</th>
<th>Student</th>
<th>HoursOfStudy</th>
<th>StudentSES</th>
<th>Pretest</th>
<th>Posttest</th>
<th>VerbalSAT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length:900</td>
<td>Length:900</td>
<td>Length:900</td>
<td>Min. : 0.00</td>
<td>Min. : -3.214232</td>
<td>Min. : 200.0</td>
<td>Min. : 268.0</td>
<td>Min. : 200.0</td>
</tr>
<tr>
<td>Class : character</td>
<td>Class : character</td>
<td>Class : character</td>
<td>1st Qu.: 1.00</td>
<td>1st Qu.: -0.661919</td>
<td>1st Qu.: 360.8</td>
<td>1st Qu.: 457.0</td>
<td>1st Qu.: 307.8</td>
</tr>
<tr>
<td>Mode : character</td>
<td>Mode : character</td>
<td>Mode : character</td>
<td>Median : 3.00</td>
<td>Median : -0.008190</td>
<td>Median : 444.0</td>
<td>Median : 536.0</td>
<td>Median : 402.0</td>
</tr>
<tr>
<td>Mean : 2.52</td>
<td>Mean : -0.000978</td>
<td>Mean : 445.4</td>
<td>Mean : 534.6</td>
<td>Mean : 415.3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3rd Qu.: 4.00</td>
<td>3rd Qu.: 0.700251</td>
<td>3rd Qu.: 528.0</td>
<td>3rd Qu.: 614.0</td>
<td>3rd Qu.: 513.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max. : 5.00</td>
<td>Max. : 2.767392</td>
<td>Max. : 800.0</td>
<td>Max. : 800.0</td>
<td>Max. : 800.0</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- `nrow(allschools)` is 900 – all three combined
**bind_rows(): Results**

- Resulting dataframe:

<table>
<thead>
<tr>
<th>School</th>
<th>Classroom</th>
<th>Student</th>
<th>HoursOfStudy</th>
<th>StudentSES</th>
<th>Pretest</th>
<th>Posttest</th>
<th>VerbalSAT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length:900</td>
<td>Length:900</td>
<td>Length:900</td>
<td>Min.: 0.00</td>
<td>Min.: -3.214232</td>
<td>Min.: 200.0</td>
<td>Min.: 268.0</td>
<td>Min.: 200.0</td>
</tr>
<tr>
<td>1st Qu.:1.00</td>
<td>1st Qu.: -0.661919</td>
<td>1st Qu.:360.8</td>
<td>1st Qu.: 457.0</td>
<td>1st Qu.: 307.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median : 3.00</td>
<td>Median : -0.008190</td>
<td>Median :440.0</td>
<td>Median :536.0</td>
<td>Median :402.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean : 2.52</td>
<td>Mean : -0.000978</td>
<td>Mean :445.4</td>
<td>Mean :534.6</td>
<td>Mean :415.3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3rd Qu.:4.00</td>
<td>3rd Qu.: 0.700251</td>
<td>3rd Qu.: 528.0</td>
<td>3rd Qu.: 614.0</td>
<td>3rd Qu.: 513.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max.: 5.00</td>
<td>Max.: 2.767392</td>
<td>Max.: 800.0</td>
<td>Max.: 800.0</td>
<td>Max.: 800.0</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **bind_rows()** is *smart*!
  - Not a problem that column order varies across dataframes
    - Looks at the column names
  - Not a problem that *VerbalSAT* column only existed in one of the original dataframes
    - **NA** (missing data) for the students at the other schools
bind_rows(): Results

- Resulting dataframe:

![Dataframe Table]

- The one “gotcha”: Factor variables
  - Remember that a factor is a variable with a **fixed set of categories**
  - When the set of categories **differ** across the dataframes, `bind_rows()` changes to a character
  - Simple solution is just to change back to a factor
    - `allschools$School <- as.factor(allschools$School)`
**bind_rows(): Results**

- Resulting dataframe:

```
<table>
<thead>
<tr>
<th>School</th>
<th>Classroom</th>
<th>Student</th>
<th>HoursOfStudy</th>
<th>StudentSES</th>
<th>Pretest</th>
<th>Posttest</th>
<th>VerbalSAT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Length:900</td>
<td>Length:900</td>
<td>Length:900</td>
<td>Min.:0.00</td>
<td>Min.:200.0</td>
<td>Min.:268.0</td>
<td>Min.:200.0</td>
</tr>
<tr>
<td>Class:character</td>
<td>Class:character</td>
<td>Class:character</td>
<td>1st Qu.:1.00</td>
<td>1st Qu.:360.8</td>
<td>1st Qu.:457.0</td>
<td>1st Qu.:307.8</td>
<td></td>
</tr>
<tr>
<td>Mode:character</td>
<td>Mode:character</td>
<td>Mode:character</td>
<td>Median:3.00</td>
<td>Median:444.0</td>
<td>Median:536.0</td>
<td>Median:402.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mean:2.52</td>
<td>Mean:-0.00978</td>
<td>Mean:445.4</td>
<td>Mean:534.6</td>
<td>Mean:415.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3rd Qu.:4.00</td>
<td>3rd Qu.:0.700251</td>
<td>3rd Qu.:528.0</td>
<td>3rd Qu.:614.0</td>
<td>3rd Qu.:513.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Max.:5.00</td>
<td>Max.:2.767392</td>
<td>Max.:800.0</td>
<td>Max.:800.0</td>
<td>Max.:800.0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
```

- You can also add the optional `.id` argument
  - `allschools <- bind_rows(school1, school2, school3, .id='OriginalDataframe')`
  - Adds another column that tracks which of the original dataframes (by number) each observation came from
**bind_rows() vs. union()**

- **bind_rows()** pastes together every row from every dataframe, even if there are duplicates
  - If you want to skip duplicates, use **union()**
    - Same syntax as **bind_rows()**, just different function name

- Other related functions:
  - **intersect()**: Keep only the rows that appear in *all* of the source dataframes
  - **setdiff()**: Keep only the rows that appear in a *single* source dataframe—if duplicates, delete both copies
Week 13: Data Management & Level-2 Variables

- Finish Power
  - Your Own Power Analysis
  - Influences on Power
- Data Management in R
  - Combining More of the Same
    - Joins
    - Reshaping Data
- Level-2 Fixed & Random Effects
  - Understanding Level-2 Variables
    - Level-2 Random Effects
    - BLUPs
    - Level-2 Fixed Effects
  - Continuous or categorical?
    - Median splits
    - Extreme groups design
  - Good measurement
    - Reliability
    - Validity
Joins

- 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

<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</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</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</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</td>
<td>C001</td>
<td>S0004</td>
<td>3</td>
<td>1.6873593</td>
<td>0.3640230</td>
<td>0.6022264</td>
</tr>
<tr>
<td>5</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</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>

**allschools**: 1 row per **student**

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

- 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.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Word</th>
<th>PrevTrials</th>
<th>RT</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>panther</td>
<td></td>
<td>0.703.877</td>
</tr>
<tr>
<td>S1</td>
<td>drive</td>
<td></td>
<td>1.532.387</td>
</tr>
<tr>
<td>S1</td>
<td>monorail</td>
<td></td>
<td>2.731.882</td>
</tr>
<tr>
<td>S13</td>
<td>peony</td>
<td></td>
<td>0.808.392</td>
</tr>
<tr>
<td>S13</td>
<td>monorail</td>
<td></td>
<td>1.489.479</td>
</tr>
<tr>
<td>S13</td>
<td>aardvark</td>
<td></td>
<td>2.875.799</td>
</tr>
</tbody>
</table>

**lexicaldecision.csv**: 1 row per **trial**

Each word appears in multiple rows.

**subtlexus.csv**: Each word has only **one** row with its frequency.
Joins

- 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

<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>

1 row per **trial**

Each subject has **multiple rows**

Each subject has only **one** row with his or her Reading Span score
Joins

- “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

**lexicaldecision.csv:**

- 1 row per trial
- Each word appears in multiple rows

**subtlexus.csv:**

- Each word has only one row with its frequency
Inner Join

- `lexdec2 <- inner_join(lexicaldecision, subtlexus, by='Word')`
  - New dataframe has both the columns from `lexicaldecision` (Subject, PrevTrials, RT) and the columns from `subtlexus` (WordFreq)
  - Matches the observations using the `Word` column
**Inner Join**

- `lexdec2 <- inner_join(lexicaldecision, subtlexus, by='Word')`
  - New dataframe has both the columns from `lexicaldecision` (Subject, PrevTrials, RT) and the columns from `subtlexus` (WordFreq)
  - Matches the observations using the **Word** column
  - Again, need to turn character variables back into factors
  - 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 `lexicaldecision` tells us which *Word* to look for in `subtlexus` ... but R doesn’t know that!
  - `lexdec2 <- inner_join(lexicaldecision, subtlexus, by=c('Item'='Word'))`
Other Types of Joins

- `nrow(lexicaldecision)` 2040
- `nrow(lexdec2)` 1800
- Six words don’t have a frequency measurement
- An `inner join` will drop rows that can’t be matched
- Alternative:
  - `lexdec2 <- left_join(lexicaldecision, subtlexus, by='Word')`

Keep the rows in the first dataframe (`lexicaldecision`) 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

- `nrow(lexicaldecision)` 2040
- `nrow(lexdec2)` 1800
- Six words don’t have a frequency measurement
- An `inner join` will drop rows that can’t be matched
- Alternative:
  - `lexdec2 <- left_join(lexicaldecision, subtlexus, by='Word')`

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

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

- `nrow(lexicaldecision)` 2040
- `nrow(lexdec2)` 1800

- Six words don’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
  - `lexdec3 <- full_join(lexicaldecision, subtlexus, by='Word')`
  - Includes rows for all of the words 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:
  
  ```r
  lexdec2 <- inner_join(lexicaldecision, subtlexus, by=c('Word', 'Country'))
  ```

  - 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.
Joins – Practice!

- Remember our math tutoring data?:

Use the functions we just discussed to add the tutor data from tutoruse to allschools.
Joins – Practice!

- Remember our math tutoring data?:

Use the functions we just discussed to add the tutor data from tutoruse to allschools:

- `allschools <- left_join(allschools, tutoruse, by=c('Classroom'='Class'))`
- `inner_join()` could also be used here
  - No missing data for any of the classrooms, so these will produce identical results
Week 13: Data Management & Level-2 Variables

- Finish Power
  - Your Own Power Analysis
  - Influences on Power
- Data Management in R
  - Combining More of the Same
  - Joins
    - Reshaping Data
- Level-2 Fixed & Random Effects
  - Understanding Level-2 Variables
    - Level-2 Random Effects
    - BLUPs
    - Level-2 Fixed Effects
  - Continuous or categorical?
    - Median splits
    - Extreme groups design
  - Good measurement
    - Reliability
    - Validity
Long vs. Wide Format

- For \texttt{lmer()}\texttt{()}, each observation needs its own row
  - “long” format

- Sometimes data comes to us in “wide” format
  - Each repeated measure is a different \textit{column} in the same row
  - e.g., what SPSS uses

<table>
<thead>
<tr>
<th>SUBJECT</th>
<th>SESSION</th>
<th>MATHSCORE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Pre-test</td>
<td>-1.27</td>
</tr>
<tr>
<td>1</td>
<td>Post-test</td>
<td>0.33</td>
</tr>
<tr>
<td>2</td>
<td>Pre-test</td>
<td>0.25</td>
</tr>
<tr>
<td>2</td>
<td>Post-test</td>
<td>1.11</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Student</th>
<th>Pretest</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>S0001</td>
<td>0.5137800</td>
</tr>
<tr>
<td>2</td>
<td>S0002</td>
<td>0.2634907</td>
</tr>
</tbody>
</table>


gather()  

- `gather()` turns “wide” data into “long” data  
- Like gathering leaves in the yard  

Leaves scattered horizontally ("wide format")  
Leaves stacked vertically in a pile ("long format")


**gather()**

- `gather()` turns “wide” data into “long” data

- `allschools.gathered <- gather(allschools, measure.vars=c('Pretest', 'Posttest'))`

  - **Pretest** and **Posttest** are the columns that we want to convert into separate observations
  - (often, repeated measures on the same individual)
gather() : The Results

- `head(allschools.gathered)`

Now we have 2 rows per student: A “Pretest” row and a “Posttest” row
- `nrow(allschools)`
- `nrow(allschools.gathered)`
gather(): Naming Columns

- head(allschools.gathered)

By default, R saves the original column name in a variable called "key", and the DV in a variable called "value"
- Not very informative
- We can specify better names when calling `gather()`: `allschools.gathered <- gather(allschools, measure.vars=c('Pretest', 'Posttest'), key='Session', value='MathSAT')`
**gather()**: Naming Columns

- `head(allschools.gathered)`

<table>
<thead>
<tr>
<th>OriginalDataframe</th>
<th>School</th>
<th>Classroom</th>
<th>Student</th>
<th>HoursOfStudy</th>
<th>StudentSES</th>
<th>VerbalSAT</th>
<th>Session</th>
<th>MathSAT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>Jefferson</td>
<td>C001</td>
<td>S0001</td>
<td>1</td>
<td>0.7803573</td>
<td>Pretest</td>
<td>632</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>Jefferson</td>
<td>C001</td>
<td>S0002</td>
<td>1</td>
<td>-0.4936052</td>
<td>Pretest</td>
<td>470</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>Jefferson</td>
<td>C001</td>
<td>S0003</td>
<td>0</td>
<td>-2.0015550</td>
<td>Pretest</td>
<td>667</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>Jefferson</td>
<td>C001</td>
<td>S0004</td>
<td>4</td>
<td>-0.9005014</td>
<td>Pretest</td>
<td>475</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>Jefferson</td>
<td>C001</td>
<td>S0005</td>
<td>4</td>
<td>-0.8569605</td>
<td>Pretest</td>
<td>698</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>Jefferson</td>
<td>C001</td>
<td>S0006</td>
<td>0</td>
<td>0.2904948</td>
<td>Pretest</td>
<td>800</td>
</tr>
</tbody>
</table>

- By default, R saves the original column name in a variable called “key”, and the DV in a variable called “value”
  - Not very informative
  - We can specify better names when calling `gather()`: `allschools.gathered <- gather(allschools, measure.vars=c('Pretest', 'Posttest'), key='Session', value='MathSAT')`
**gather(): Naming Columns**

- `head(allschools.gathered)`

- Now, you can run an `lmer()` model where MathSAT is the DV and `Session` is one of the predictor variables.
**Reshaping Data**

- `gather()` turns “wide” data into “long” data

- Also a corresponding function, `spread()`, to turn “long” format data into “wide” format data
Week 13: Data Management & Level-2 Variables

- Finish Power
  - Your Own Power Analysis
  - Influences on Power
- Data Management in R
  - Combining More of the Same
  - Joins
  - Reshaping Data
- Level-2 Fixed & Random Effects
  - Understanding Level-2 Variables
    - Level-2 Random Effects
    - BLUPs
    - Level-2 Fixed Effects
  - Continuous or categorical?
    - Median splits
    - Extreme groups design
  - Good measurement
    - Reliability
    - Validity
Summary

- Same variables in more than one file:
  - `bind_rows()`

- Different variables in more than one file:
  - `inner_join()`, `left_join()`

- Data is already in one data frame but you need to rearrange it:
  - `gather()`
Agnieszka is studying the relationship between personality & moral judgments (like the “trolley problem”). In the dataframe `trials`, each row is one Subject’s response to one moral dilemma (out of 40 total). A separate dataframe `personality` holds the results from a personality inventory, where each Subject has 1 row with their results on the Big 5 personality scales.
Agnieszka is studying the relationship between personality & moral judgments (like the “trolley problem”). In the dataframe `trials`, each row is one Subject’s response to one moral dilemma (out of 40 total). A separate dataframe `personality` holds the results from a personality inventory, where each Subject has 1 row with their results on the Big 5 personality scales.

```r
inner_join(trials, personality, by='Subject')
```
Data Management Practice

- We conduct a longitudinal study in which we assess high schoolers’ social networks in each year of high school. Because the assessments were done in different years, the data from each year is saved on four different computers, in dataframes named firstyear, sophomore, junior, and senior.
We conduct a longitudinal study in which we assess high schoolers’ social networks in each year of high school. Because the assessments were done in different years, the data from each year is saved on four different computers, in dataframes named `firstyear`, `sophomore`, `junior`, and `senior`.

```
bind_rows(firstyear, sophomore, junior, senior)
```
Data Management Practice

- We are re-analyzing some existing data on brainstorming ideas for new technology. Each row of the dataframe `brainstorming` consists of the data for one Subject, with one column indicating the number of ideas brainstormed in the Individual condition and another column the number of ideas in the Group condition. We want to be able to re-analyze the data using a mixed-effects model.
Data Management Practice

- We are re-analyzing some existing data on brainstorming ideas for new technology. Each row of the dataframe `brainstorming` consists of the data for one `Subject`, with one column indicating the number of ideas brainstormed in the `Individual` condition and another column the number of ideas in the `Group` condition. We want to be able to re-analyze the data using a mixed-effects model.

  ```r
gather(brainstorming, measure.vars=c('Individual', 'Group'))
```
Week 13: Data Management & Level-2 Variables

- Finish Power
  - Your Own Power Analysis
  - Influences on Power
- Data Management in R
  - Combining More of the Same
  - Joins
- Reshaping Data
- Level-2 Fixed & Random Effects
  - Understanding Level-2 Variables
    - Level-2 Random Effects
    - BLUPs
    - Level-2 Fixed Effects
  - Continuous or categorical?
    - Median splits
    - Extreme groups design
  - Good measurement
    - Reliability
    - Validity
**Level-2 Random Effects**

- Let’s consider one model of lexical decision:
  - Center `PrevTrials`, then include it in a model:

  ```r
  model1 <- lmer(RT ~ 1 + PrevTrials.cen + (1+PrevTrials.cen||Subject) + (1|Word),
  data=lexdec2)
  ```
Let’s consider one model of lexical decision:

- Center `PrevTrials`, then include it in a model:
  - `lexdec2$PrevTrials.cen <- scale(lexdec2$PrevTrials, center=TRUE, scale=FALSE)[,1]`
  - or
  - `lexdec2$PrevTrials.cen <- lexdec2$PrevTrials - mean(lexdec2$PrevTrials)`

- `model1 <- lmer(RT ~ 1 + PrevTrials.cen + (1+PrevTrials.cen||Subject) + (1|Word), data=lexdec2)`
Level-2 Random Effects

- Let’s consider one model of lexical decision:
  - `model1 <- lmer(RT ~ 1 + PrevTrials.cen + (1+PrevTrials.cen||Subject) + (1|Word), data=lexdec2)`

What do these fixed and random intercepts mean, really?
Level-2 Random Effects

- Think back to a normal distribution…
  - The *standard normal* has *mean 0* and *standard deviation 1*
Level-2 Random Effects

- We can also have normal distributions with other means and standard deviations
  - This one has mean 40 and standard deviation 6
Level-2 Random Effects

- Back to our model results...
  - Fixed intercept is the mean RT averaging across all subjects: 604 ms
  - But some subjects have faster RTs than others
    - People read at different speeds! (duh)
    - So, we have a *distribution* of reading speeds
  - The SD of reading speeds is 73 ms
Level-2 Random Effects

- There is a distribution of people’s average lexical decision times
  - The average person has an RT of 604 ms (fixed intercept)
  - But, there is a standard deviation of 73 ms (random intercept)

64% of people have an average RT between 531 and 677 ms
**Level-2 Random Effects**

- Our subjects are a random sample from this population of people with different average RTs.
- Assumed in `lmer()` and `glmer()` to be a normal distribution.

64% of people have an average RT between 531 and 677 ms.
Level-2 Random Effects

- Another example: IQ scores have a well-defined distribution
  - Average person has IQ score of 100 (fixed intercept)
  - The distribution of IQ scores has a standard deviation of 20 (random intercept)

64% of people have an IQ between 80 and 120
Level-2 Random Effects

- We can apply the same reasoning to the PrevTrials effect
  - For the average person, RTs decrease by 6 ms with every additional trial of experience (fixed effect)
  - But, some people learn at faster or slower rates
    - So, we have a *distribution* of PrevTrials effects—some people have relatively fast learning rates and some have relatively slow
  - The SD of this distribution is 3 ms/trial

<table>
<thead>
<tr>
<th>Random effects:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Groups</td>
</tr>
<tr>
<td>--------</td>
</tr>
<tr>
<td>Word</td>
</tr>
<tr>
<td>Subject</td>
</tr>
<tr>
<td>Subject.1</td>
</tr>
<tr>
<td>Residual</td>
</tr>
<tr>
<td>Number of obs: 1800, groups: Word, 45; Subject, 40</td>
</tr>
</tbody>
</table>

Fixed effects:

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>602.556</td>
<td>13.4601</td>
<td>44.81</td>
</tr>
<tr>
<td>PrevTrials.cen</td>
<td>-6.3853</td>
<td>0.6054</td>
<td>-10.55</td>
</tr>
</tbody>
</table>
Level-2 Random Effects

- A *distribution* of PrevTrials effects (i.e., of learning rates) in the population
  - The **average** person has an effect of 6 ms/trial (fixed slope)
  - But, there is a **standard deviation** of 3 ms/trial (random slope)

64% of people have an PrevTrials effect between 3 and 9 ms
Week 13: Data Management & Level-2 Variables

- Finish Power
  - Your Own Power Analysis
  - Influences on Power
- Data Management in R
  - Combining More of the Same
  - Joins
  - Reshaping Data
- Level-2 Fixed & Random Effects
  - Understanding Level-2 Variables
    - Level 2 Random Effects
    - BLUPs
    - Level-2 Fixed Effects
- Continuous or categorical?
  - Median splits
  - Extreme groups design
- Good measurement
  - Reliability
  - Validity
### BLUPs

- Where do individual subjects fall in this distribution?
  - `ranef(model1)`
    - Shows you the intercepts and slopes for individual subjects & items
    - These are *adjustments* relative to the *fixed effect*
    - **Best Linear Unbiased Predictors (BLUPs)**

<table>
<thead>
<tr>
<th>Subject</th>
<th>PrevTrials.cen</th>
<th>(Intercept)</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>-2.57223423</td>
<td>-24.215052</td>
</tr>
<tr>
<td>S10</td>
<td>1.08186983</td>
<td>28.833723</td>
</tr>
<tr>
<td>S11</td>
<td>-1.62941789</td>
<td>-39.140742</td>
</tr>
<tr>
<td>S12</td>
<td>-3.77777882</td>
<td>-118.154257</td>
</tr>
<tr>
<td>S13</td>
<td>-1.16062561</td>
<td>-56.087963</td>
</tr>
<tr>
<td>S14</td>
<td>0.39431901</td>
<td>-60.570369</td>
</tr>
<tr>
<td>S15</td>
<td>1.25693444</td>
<td>23.731177</td>
</tr>
<tr>
<td>S16</td>
<td>-1.04021645</td>
<td>-84.679952</td>
</tr>
<tr>
<td>S17</td>
<td>1.79180123</td>
<td>26.226645</td>
</tr>
<tr>
<td>S18</td>
<td>-0.94302648</td>
<td>-11.473677</td>
</tr>
</tbody>
</table>

Example: Subject S12 has a mean RT 118 ms lower than the average. (Fast responder!)

Average is **603 ms**, so this person’s average RT is:  
\[603 - 118 = 485 \text{ ms}\]
BLUPs

- Where do individual subjects fall in this distribution?
  - `ranef(model1)`
    - Shows you the intercepts and slopes for individual subjects & items
    - These are *adjustments* relative to the fixed effect
    - **Best Linear Unbiased Predictors (BLUPs)**

<table>
<thead>
<tr>
<th>Subject</th>
<th>PrevTrials.cen</th>
<th>(Intercept)</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>-2.57223423</td>
<td>-24.215052</td>
</tr>
<tr>
<td>S10</td>
<td>1.08186983</td>
<td>28.833723</td>
</tr>
<tr>
<td>S11</td>
<td>-1.62941789</td>
<td>-39.140742</td>
</tr>
<tr>
<td><strong>S12</strong></td>
<td>-3.77777882</td>
<td>-118.154257</td>
</tr>
<tr>
<td>S13</td>
<td>-1.16062561</td>
<td>-56.087963</td>
</tr>
<tr>
<td>S14</td>
<td>0.39431901</td>
<td>-60.570369</td>
</tr>
<tr>
<td>S15</td>
<td>1.25693444</td>
<td>23.731177</td>
</tr>
<tr>
<td>S16</td>
<td>-1.04021645</td>
<td>-84.679952</td>
</tr>
<tr>
<td>S17</td>
<td>1.79180123</td>
<td>26.226645</td>
</tr>
<tr>
<td>S18</td>
<td>-0.94302648</td>
<td>-11.473670</td>
</tr>
</tbody>
</table>

S12 also a PrevTrials slope that is -4 relative to the overall fixed effect.

Fixed effect of PrevTrials is -6 ms, so this person’s PrevTrials effect is:
-6 – -4 = -10 ms/trial (stronger effect)
BLUPs

- Why do you think the BLUPs aren’t displayed in our initial results from `summary()`?
  - For random effects, we’re mainly interested in modeling variability
  - BLUPs aren’t considered parameters of the model
    - Not what this is a model “of”
    - We ran this analysis to model the effects of experience on lexical decision, not the effect of being Dave from Altoona
  - If we ran the same design with a different sample, BLUPs probably wouldn’t be the same
    - No reason to expect that Subject 12 in the new sample will again be one of the faster subjects
    - By contrast, we do intend for our fixed effects to replicate
**BLUPs**

- Can you figure out which **Item** has the *fastest* RT relative to the mean? Which has the slowest:
  - Fastest: *panther* (-42 ms vs. mean)
  - Slowest: *clogs* (+43 ms vs. mean)
Week 13: Data Management & Level-2 Variables

- Finish Power
  - Your Own Power Analysis
  - Influences on Power
- Data Management in R
  - Combining More of the Same
  - Joins
  - Reshaping Data
- Level-2 Fixed & Random Effects
  - Understanding Level-2 Variables
    - Level 2 Random Effects
    - BLUPs
    - Level-2 Fixed Effects
  - Continuous or categorical?
    - Median splits
    - Extreme groups design
- Good measurement
  - Reliability
  - Validity
Level-2 Fixed Effects

- Let’s consider another model of our lexical decision data:
  - model2 <- lmer(RT ~ 1 + PrevTrials.cen + (1|Subject) + (1|Word), data=lexdec2)

- Hierarchical linear model notation for this:
  - Lv.2 (Item): $B_k = u_{00(k)}$
  - Lv.2(Subj.): $B_j = u_{00(j)}$
  - Lv.1(Trial): $Y_{i(jk)} = \gamma_{000} + \gamma_{100}\text{PrevTrials} + B_j + B_k + \epsilon_{i(jk)}$

Level 2 model predicts the effect of item $k$  
Could substitute random intercept into the level 1 model
Level-2 Fixed Effects

- Now let’s add a fixed effect of word frequency:
  - `model3 <- lmer(RT ~ 1 + PrevTrials.cen + WordFreq + (1|Subject) + (1|Word), data=lexdec2)`

- Which level does this characterize?:
  - Lv.2 (Item): $B_k = u_{00(0k)}$
  - Lv.2(Subj.): $B_j = u_{00(j0)}$
  - Lv.1(Trial): $Y_{ijk} = \gamma_{000} + \gamma_{100}\text{PrevTrials} + B_j + B_k + e_{ijk}$

  Level 2 model predicts the effect of item $k$

  Could substitute random intercept into the level 1 model
**Level-2 Fixed Effects**

- Now let’s add a fixed effect of word frequency:
  - `model3 <- lmer(RT ~ 1 + PrevTrials.cen + WordFreq + (1|Subject) + (1|Word),
                     data=lexdec2)`

- Which level does this characterize?:
  - **Lv.2 (Item):** $B_k = \gamma_{200} \text{WordFreq} + u_{00(0k)}$
  - **Lv.2 (Subj.):** $B_j = u_{00(j0)}$
  - **Lv.1 (Trial):** $Y_{i(jk)} = \gamma_{000} + \gamma_{100} \text{PrevTrials} + B_j + B_k + e_{i(jk)}$
    - Intercept
    - # of previous trials seen
    - Subject
    - Item
    - Error
What Changes?

• **Random** item variance is greatly reduced.
• Word frequency accounts for a lot of the variance among items.
• Word frequency explains a lot of the “Item $k$” effect we’re substituting into the level 1 equation. No longer just a random intercept.
What Didn’t Change?

- Level 1 fixed effect (PrevTrials) and error term essentially unchanged.
- Doesn’t matter what explains the “Item k” effect; still substituting into the same Lv 1 model
- Note that WordFreq & PrevTrials effects are slightly correlated (due to random sampling of item orders); otherwise, there’d be no change.
What Didn’t Change?

- Estimated variance in subject intercept also essentially the same
- Explaining where the “Item k” effect comes from doesn’t change the “Subject j” effect
Week 13: Data Management & Level-2 Variables

- Finish Power
  - Your Own Power Analysis
  - Influences on Power
- Data Management in R
  - Combining More of the Same
  - Joins
  - Reshaping Data
- Level-2 Fixed & Random Effects
  - Understanding Level-2 Variables
    - Level 2 Random Effects
    - BLUPs
    - Level-2 Fixed Effects
  - Continuous or categorical?
    - Median splits
    - Extreme groups design
  - Good measurement
    - Reliability
    - Validity
Continuous or Categorical Predictors?

- In ANOVA, subject & item differences typically examined as **categorical** variables
- e.g. **median split**:
  - \( \text{median(loxdec2$WordFreq, na.rm=TRUE)} = 3.30 \)
  - Word frequencies above the median are in category A and words below it are in category B

![Graph showing word frequency vs mean RT](image-url)
Continuous or Categorical Predictors?

- In ANOVA, subject & item differences typically examined as categorical variables
- e.g. median split:
  - median(lexcdec2$WordFreq, na.rm=TRUE) = 3.30
  - Word frequencies above the median are in category A and words below it are in category B
Evaluating Median Splits

- Median splits are noisy and discard info.
- Ignores all within-category variation

Median split considers these both equally “low-frequency” words

- pomegranate (WF: 1.1461)
- glasses (WF: 3.2279)
Evaluating Median Splits

- Median splits are **noisy** and **discard info**.
  - Ignores all within-category variation
  - High probability of misclassification

If our measures of word frequency were even slightly off, these words could have ended up in the opposite categories!

- glasses (WF: 3.2729)
- chair (WF: 3.400)
Evaluating Median Splits

- Median splits are **noisy** and **discard info.**
  - Ignores all within-category variation
  - High probability of misclassification
- Greatly reduces power and estimated effect size (Cohen, 1983)
Median splits are **noisy** and **discard info.**
- Ignores all within-category variation
- High probability of misclassification

Greatly reduces power and estimated effect size (Cohen, 1983)

Also, comparing two categories can’t tell us about the form of the relationship (as polynomial contrasts can)

If continuous variation (in word frequency, second language proficiency, etc.) measured, better to include it in the model
**Week 13: Data Management & Level-2 Variables**

- Finish Power
  - Your Own Power Analysis
  - Influences on Power
- Data Management in R
  - Combining More of the Same
  - Joins
  - Reshaping Data
- Level-2 Fixed & Random Effects
  - Understanding Level-2 Variables
    - Level 2 Random Effects
    - BLUPs
    - Level-2 Fixed Effects
  - Continuous or categorical?
    - Median splits
    - Extreme groups design
  - Good measurement
    - Reliability
    - Validity
Extreme Group Designs

- In some cases, we might deliberately sample only very low- and very high-frequency words
  - Extreme group design
- Now, we don’t know what the full relation is

Here there be dragons
Extreme Group Designs

- In some cases, we might deliberately sample only very low- and very high-frequency words
  - Extreme group design
- Now, we don’t know what the full relation is
Extreme Group Designs

- In some cases, we might deliberately sample only very low- and very high-frequency words
  - Extreme group design
- Now, we don’t know what the full relation is

![Extreme Group Designs](chart.png)
In some cases, we might deliberately sample only very low- and very high-frequency words.

- Extreme group design

Now, we don’t know what the full relation is.

- Should treat this as a categorical variable (reflects design)
Extreme Group Designs: Evaluation

- May overestimate effect size
- Still, better than median splits if you want to do a categorical design (Conway et al., 2005)
  - e.g., you only care whether a difference exists (not its size / shape)
**Breakpoints**

- When you have a continuous variable, but you think there’s a qualitative shift at some point in the range
  - e.g., below vs. above the poverty line
- Add a categorical variable that represents whether or not you’re above the point at which the shift happens

**Main effect of breakpoint only** – single shift downward but same slope

**Main effect of breakpoint & an interaction** – slope also changes
Week 13: Data Management & Level-2 Variables

- Finish Power
  - Your Own Power Analysis
  - Influences on Power
- Data Management in R
  - Combining More of the Same
  - Joins
  - Reshaping Data
- Level-2 Fixed & Random Effects
  - Understanding Level-2 Variables
    - Level 2 Random Effects
    - BLUPs
    - Level-2 Fixed Effects
  - Continuous or categorical?
    - Median splits
    - Extreme groups design
  - Good measurement
    - Reliability
    - Validity
Good Measurement: Reliability

- Suppose we find that a measure of working memory is unrelated to people’s moral judgments
  - Maybe these are truly unrelated
  - Or, maybe we just failed to accurately measure WM and/or moral reasoning

- Not all measures are good measures
  - Measures may be noisy
  - Measures may not measure a stable or meaningful characteristic of people/items/schools
Good Measurement: Reliability

- Good measures produce consistent scores
  - Across times (test-retest reliability)
  - Across items (internal consistency)
  - Across judges (inter-rater reliability)
- Shows you’re measuring something real

- If measures can’t even predict themselves, they can’t predict anything else!

$r = .77$ Good!

$r = .16$ Bad!

e.g., James, Fraundorf, Lee, & Watson, 2018
Week 13: Data Management & Level-2 Variables

- Finish Power
  - Your Own Power Analysis
  - Influences on Power
- Data Management in R
  - Combining More of the Same
  - Joins
  - Reshaping Data
- Level-2 Fixed & Random Effects
  - Understanding Level-2 Variables
    - Level 2 Random Effects
    - BLUPs
    - Level-2 Fixed Effects
  - Continuous or categorical?
    - Median splits
    - Extreme groups design
  - Good measurement
    - Reliability
    - Validity
Good Measurement: Validity

- Even if we have a reliable measure, no guarantee it measures the thing we think it measures
  - You’re measuring something, but what is it?
  - *Examples of tests that produce consistent results but don’t measure what we want:*

  - Stopped clock
  - Polygraph

<table>
<thead>
<tr>
<th>BMI Chart</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>BMI less than 18.50</strong></td>
</tr>
<tr>
<td><strong>BMI 18.50 - 24.99</strong></td>
</tr>
<tr>
<td><strong>BMI 25.00 - 29.99</strong></td>
</tr>
<tr>
<td><strong>BMI 30 or more</strong></td>
</tr>
</tbody>
</table>
Good Measurement: Validity

- Valid measures should show (among other things):
  - **Convergent validity**: Correlate with *other* measures of this construct

Operation Span task: Remember letters while verifying equations

3 x 4 = 12  
(T / F)?

Reading Span task: Remember letters while verifying sentences

An official who manages a state is called a governor.  
(T / F)?

Here, two tasks designed to measure working memory correlate.
Good Measurement: Validity

- Valid measures should show (among other things):
  - **Convergent validity**: Correlate with *other* measures of this construct
  - **Divergent validity**: *Don’t* correlate with things that are supposed to be *different*
    - If “working memory” task correlates with years of education or socioeconomic status, might not be measuring what we thought
**Good Measurement: Validity**

- Valid measures should show (among other things):
  - **Convergent validity**: Correlate with *other* measures of this construct
  - **Divergent validity**: *Don’t* correlate with things that are supposed to be *different*
    - Do higher Working Memory scores predict second language learning just because subjects who are “smarter” or more motivated do well on both tasks?
    - Or is this unique to WM?
    - Measuring only 1 construct makes it difficult to tell where the locus of an effect lies
Week 13: Data Management & Level-2 Variables

- Finish Power
  - Your Own Power Analysis
  - Influences on Power
- Data Management in R
  - Combining More of the Same
  - Joins
  - Reshaping Data
- Level-2 Fixed & Random Effects
  - Understanding Level-2 Variables
    - Level 2 Random Effects
    - BLUPs
    - Level-2 Fixed Effects
  - Continuous or categorical?
    - Median splits
    - Extreme groups design
  - Good measurement
    - Reliability
    - Validity
That's all Folks!