Course Business

- New sample dataset for class today:
  - CourseWeb: Course Documents → Sample Data → Week 3

- How to delete a variable completely:
  - `experiment$TrialsRemaining <- NULL`

- Brown bag on data visualization in R (Kevin Soo & Cory Derringer)
  - Wednesday, January 30th, noon
  - Same location as this class
Distributed Practice!

- How can I make a comment in R?
- What R function would be best to get the mean GPA for each school in my study?
- ...to get an overall look at the means of each variable?
Distributed Practice!

- How can I make a comment in R? #
- What R function would be best to get the mean GPA for each school in my study?
  - `tapply()`
- …to get an overall look at the means of each variable?
  - `summary()`

---

The Spacing Effect

*A Case Study in the Failure to Apply the Results of Psychological Research*

Frank N. Dempster
University of Nevada, Las Vegas

**ABSTRACT:** The spacing effect would appear to have considerable potential for improving classroom learning, yet there is no evidence of its widespread application. I consider nine possible impediments to the implementation of research findings in the classroom in an effort to determine which, if any, apply to the spacing effect. I conclude that the apparent absence of systematic application may be due, in part, to the historical character of research on the spacing effect and certain gaps in our understanding of both the spacing effect and classroom practice. However, because none of these concerns seems especially discouraging, and in view of what we do know about the spacing effect, classroom application is recommended.

The spacing effect—which refers to the finding that for a given amount of study time, spaced presentations yield substantially better learning than do massed presentations—is one of the most remarkable phenomena to emerge from laboratory research on learning. It is remarkable in several respects. First, the spacing effect is one of the most dependable and replicable phenomena in experimental psychology. Second, it is remarkably robust. In many cases, two spaced presentations are about twice as effective as two massed presentations (e.g., Hintzman, 1974; Melton, 1970), and the difference between them increases as the frequency of repetition increases (Underwood, 1970). Moreover, demonstrations of achievement following massed presentations often are only slightly higher than that following a single presentation (e.g., Melton, 1970). Third, the spacing effect is truly ubiquitous in scope. It has been observed in virtually every standard experimental learning paradigm, with all sorts of traditional research material (Dempster, 1987a; Hintzman, 1974; Melton, 1970).

With all of these characteristics in its favor, the spacing effect would seem to have considerable potential for improving classroom learning. However, there is little evidence that this potential has been realized. Neither American classrooms nor American textbooks appear to implement spaced reviews in any systematic way, and by comparison, Soviet mathematics textbooks provide a much more distributed method of presentation than do their American counterparts (Stigler, Fusan, Hsu, & Kim, 1986). Nor is there much evidence that the next generation of educators is being better informed. In a recent sampling of practitioner-oriented textbooks suitable for use in teacher education programs, I found either little or no mention of the practical benefits of the spacing effect, and in some cases the spacing effect was confused with other phenomena (e.g., Good & Brophy, 1986; Mayer, 1987; Slavin, 1986; Woodfolk, 1987). One well-known educator, in fact, advised against spaced practice at least in the early stages of learning (Hunter, 1983).

Why is it that research findings that appear to have significant implications, such as the spacing effect, often are not utilized by teachers and curriculum makers? In general, the problem is that there is no well-developed implementation model, nor is there a standard methodology for analyzing the conditions that foster the transfer of knowledge from the laboratory to the classroom (see Howland, 1984, for a discussion). Obviously, issues regarding the utilization of findings from basic research are complicated, and there are many potential impediments to the implementation of research findings in the classroom. In this article, I explore nine potential impediments, all of which seem reasonable at first glance, in an effort to determine which, if any, apply to the spacing effect.

**Impediments to Application**

*The Phenomenon Has Not Been Known Long Enough*

Although the time lag between discovery and application varies greatly, some considerable period of time often intervenes between the publication of research findings and their application. In the case of the spacing effect, however, a considerable period of time already has passed since its initial documentation. The spacing effect was known as early as 1885 when Ebbinghaus published the results of his seminal experimental work on memory. With himself as the subject, Ebbinghaus found that for a single 12-syllable series, 68 immediately successive repetitions had the effect of making possible an errorless recall after seven additional repetitions on the following day. However, the same effect was achieved by only 38 distributed repetitions spread over three days. On the basis of this and other related findings, Ebbinghaus concluded that “with any considerable number of repetitions a suitable distribution of them over a space of time is decidedly more advantageous than the massing of them at a single time” (Ebbinghaus, 1885/1913, p. 89). Just, also working with non-
Week 3: Fixed Effects

- Installing Packages
- Fixed Effects
  - Introduction to Fixed Effects
  - Running the Model in R
  - Hypothesis Testing
  - Model Formulae
  - Interpreting Interactions
  - Model Fitting
  - Fitted Values, Residuals, & Outliers
- Effect Size
  - Unstandardized
  - Standardized
  - Interpretation
  - Overall Variance Explained
R Packages

• R has lots of add-ons for many kinds of statistical analysis (e.g., structural equation modeling)

• *lme4*: Package for mixed effects models
Downloading the Package: RStudio

- **Tools** menu -> **Install Packages**...
- Type in `lme4`
- Leave **Install Dependencies** checked
  - Grabs the other packages that `lme4` makes use of
  - Only need to do this once per computer!
**Downloading the Package: R**

- **Packages & Data menu** -> **Package Installer** -> **Get List**
- **Find `lme4`**
- **Make sure to check Install Dependencies**
  - Grabs the other packages that `lme4` makes use of
  - Only need to do this once per computer!
Analyses & Add-On Packages

- Some other relevant packages:
  - **sem**: Structural equation modeling
  - **mice**: Multiple imputation of missing data
  - **psych**: Psychometrics (scale construction, etc.)
  - **party**: Random forests
  - **ggplot2**: Fancy plotting functions
  - **stringr**: Working with character variables
  - **dplyr**: Data processing, manipulation, formatting
library() command

- Need to do this in each script where you’ll use the package:
  - `library(lme4)`

- Tells R to load up the lme4 package you downloaded
  - If you had a lot of add-on packages, loading them all automatically would make R really slow to start
  - So, we only load the packages needed for this analysis
Week 3: Fixed Effects

- Installing Packages
- Fixed Effects
  - Introduction to Fixed Effects
  - Running the Model in R
  - Hypothesis Testing
  - Model Formulae
  - Interpreting Interactions
  - Model Fitting
  - Fitted Values, Residuals, & Outliers
- Effect Size
  - Unstandardized
  - Standardized
  - Interpretation
  - Overall Variance Explained
Mixed Effects Models!

- Next 3 weeks: Basics of a mixed effects analysis with continuous/numerical variables
  - This week: Fixed effects (effects of interest)
  - Next 2 weeks: Random effects (e.g., subjects, classrooms, items, firms, dyads/couples)

- After that: categorical variables
  - As predictors
  - As outcomes
Introduction to Fixed Effects

- Course Documents → Sample Data → Week 3
- Stroop task dataset
  - `Stroop <- read.csv( ... )`
Introduction to Fixed Effects

• Course Documents ➔ Sample Data ➔ Week 3
• Stroop task dataset
  • Stroop <- read.csv( ... )

green
Introduction to Fixed Effects

- Course Documents → Sample Data → Week 3
- Stroop task dataset
  - `Stroop <- read.csv( ... )`

yellow
Introduction to Fixed Effects

• Course Documents → Sample Data → Week 3
• Stroop task dataset
  • `Stroop <- read.csv( ... )`
Introduction to Fixed Effects

- Predicting one variable as a function of others
**Introduction to Fixed Effects**

- Predicting one variable as a function of others

\[ \text{Latency to name color} = \text{Baseline} + \# \text{ of previous trials} + \text{Font Size} \]

- NEXT WEEK!
Introduction to Fixed Effects

- Predicting one variable as a function of others

\[
\text{Latency to name color} = \text{Baseline} + \text{# of previous trials} + \text{Font Size}
\]

**Fixed effects** that we’re trying to model
Introduction to Fixed Effects

• Predicting one variable as a function of others

\[ Y_{000} = \text{Baseline} + \# \text{ of previous trials} + \text{Font Size} \]

(Bryk & Raudenbush, 1992; Quene & van den Bergh, 2004, 2008)
Introduction to Fixed Effects

- Predicting one variable as a function of others

\[
Y_{000} + X_{1i(jk)} + X_{2i(jk)}
\]

Latency to name color

Baseline

# of previous trials

Font Size

(Bryk & Raudenbush, 1992; Quene & van den Bergh, 2004, 2008)
Introduction to Fixed Effects

- Predicting one variable as a function of others
- Relationship of font size to RT is probably not 1:1

\[ Y_{000} = \gamma_0 + \gamma_1(x_{1i} - jk) + \gamma_2(x_{2i} - jk) \]

Latency to name color
Baseline

Regression line relating font size to RT

(Bryk & Raudenbush, 1992; Quene & )
Introduction to Fixed Effects

• Predicting one variable as a function of others
  • $\gamma_{200} = \text{slope of line relating font size to RT}$
  • One of the fixed effects we want to find this out
    • How does font size affect response time in this task?

\[
\text{Latency to name color} = Y_{000} + Y_{100}x_{1(ijk)} + Y_{200}x_{2(ijk)}
\]

Introduction to Fixed Effects

- Can we determine the exact RT based on number of previous trials & font size?
  - Probably not.
  - These variables just provide our best guess

\[ Y_{000} + Y_{100}X_{1i(jk)} + Y_{200}X_{2i(jk)} \]

Latency to name color
Baseline
# of previous trials
Font Size

(Bryk & Raudenbush, 1992; Quene & van den Bergh, 2004, 2008)
Introduction to Fixed Effects

• Can we determine the exact RT based on number of previous trials & font size?
  • Probably not.
  • These variables just provide our best guess
  • The expected value

\[
E(Y_{i(jk)}) = Y_{000} + Y_{100}x_{1i(jk)} + Y_{200}x_{2i(jk)}
\]

- Latency to name color
- Baseline
- # of previous trials
- Font Size

(Bryk & Raudenbush, 1992; Quene & van den Bergh, 2004, 2008)
Introduction to Fixed Effects

• To represent the actual observation, we need to add an error term
• Discrepancy between expected & actual value

\[ Y_{i(jk)} = Y_{000} + Y_{100}X_{1i(jk)} + Y_{200}X_{2i(jk)} \]

Latency to name color
Baseline
# of previous trials
Font Size
Error

(Bryk & Raudenbush, 1992; Quene & van den Bergh, 2004, 2008)
Introduction to Fixed Effects

- To represent the actual observation, we need to add an error term.
  - Discrepancy between expected & actual value.

\[ y_{i(jk)} = \gamma_000 + \gamma_{100} x_{1i(jk)} + \gamma_{200} x_{2i(jk)} + e_{i(jk)} \]

Latency to name color

Baseline

# of previous trials

Font Size

Error

(Bryk & Raudenbush, 1992; Quene & van den Bergh, 2004, 2008)
Introduction to Fixed Effects

• What if we aren’t interested in predicting specific values?
  • e.g., We want to know whether a variable matters or the size of its effect

• But: We learn this from asking whether and how an independent variable predicts the dependent variable
  • If font size significantly predicts what the RT will be, there’s a relation
Week 3: Fixed Effects

- Installing Packages
- Fixed Effects
  - Introduction to Fixed Effects
    - Running the Model in R
    - Hypothesis Testing
    - Model Formulae
    - Interpreting Interactions
    - Model Fitting
    - Fitted Values, Residuals, & Outliers
- Effect Size
  - Unstandardized
  - Standardized
  - Interpretation
  - Overall Variance Explained
Running the Model in R

Linear Mixed Effects Regression
Running the Model in R

• Time to fit our first model!
  • `model1 <- lmer(RT ~ 1 + PrevTrials + FontSize + (1|Subject) + (1|Item), data=Stroop)`

  Name of our model, like naming a dataframe
  Linear mixed effects regression (function name)
  Dependent measure comes before the ~
  Intercept (we’ll discuss this more very soon)
  Variables of interest (fixed effects)
  Random effect variables
  Name of the dataframe where your data is

• Here it is as a single line:
  • `model1 <- lmer(RT ~ 1 + PrevTrials + FontSize + (1|Subject) + (1|Item), data=Stroop)`
Running the Model in R

• Time to fit our first model!

  • `model1 <- lmer(RT ~ 1 + PrevTrials + FontSize + (1|Subject) + (1|Item), data=Stroop)`

    - Name of our model, like naming a dataframe
    - Linear mixed effects regression (function name)
    - Dependent measure comes before the ~
    - Intercept (we’ll discuss this more very soon)
    - Variables of interest (fixed effects)
    - Random effect variables
    - Name of the dataframe where your data is

• Quick note: This version of the model makes assumptions about the random effects that might not be true. We’ll deal with this in the next two weeks when we discuss random effects.
Running the Model in R

- Where are my results?
  - Just like with a dataframe, we’ve saved them in something we can view later.

- To view the model results:
  - `summary(model1)`
    - Or whatever your model name is.

- Saving the model makes it easy to compare models later or to view your results again.
Sample Model Results

- **Formula:** Variables you included
- **Data:** Dataframe you ran this model on
  - Check that these two matched what you wanted!
- **Random effects = next week!**
- **Relevant to model fitting. Will discuss soon.**
- **Number of observations, # of subjects, # of items**
- **Results for fixed effects of interest (next slide!)**
- **Correlations between effects**
  - Probably don’t need to worry about this unless correlations are very high (Friedman & Wall, 2005; Wurm & Fisicaro, 2014)
**Parameter Estimates**

- Estimates are the $y$ values from the model notation.

- Each additional trial of experience $\approx 18$ ms decrease in RT.
- 1-point increase in font size $\approx 13$ ms increase in RT.
- Intercept: Baseline RT if # of trials & font size are 0.
- Each of these effects are while holding the others constant.

- Core feature of multiple regression!!
- Don’t need to do residualization for this (Wurm & Fiscaro, 2014)
Parameter Estimates

WHERE THE @#$%^&@$ ARE MY P-VALUES!?
Week 3: Fixed Effects

- Installing Packages
- Fixed Effects
  - Introduction to Fixed Effects
  - Running the Model in R
    - Hypothesis Testing
    - Model Formulae
    - Interpreting Interactions
    - Model Fitting
    - Fitted Values, Residuals, & Outliers
- Effect Size
  - Unstandardized
  - Standardized
  - Interpretation
  - Overall Variance Explained
Hypothesis Testing—t test

• Reminder of why we do inferential statistics
• We know there’s some relationship between font size & RT in our sample

<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>968.5671</td>
<td>17.6221</td>
<td>54.96</td>
</tr>
<tr>
<td>PrevTrials</td>
<td>0.8583</td>
<td>-20.93</td>
<td></td>
</tr>
<tr>
<td>FontSize</td>
<td>12.7588</td>
<td>0.2309</td>
<td>55.26</td>
</tr>
</tbody>
</table>

• But:
  • Would this hold true for all people (the population) doing the Stroop?
  • Or is this sampling error? (i.e., random chance)
Hypothesis Testing—t test

- Font size effect in our sample estimated to be 12.7588 ms ... is this good evidence of an effect in the population?
- Would want to compare relative to a measure of sampling error

\[ t = \frac{\text{Estimate}}{\text{Std. error}} = \frac{12.7588}{0.2309} \]
Hypothesis Testing—t test

- We don’t have \( p \)-values (yet), but do we have a \( t \) statistic
  - Effect divided by its standard error (as with any \( t \) statistic)

- A \( t \) test comparing this \( \gamma \) estimate to 0
  - 0 is the \( \gamma \) expected under the null hypothesis that this variable has no effect

### Fixed effects:

<table>
<thead>
<tr>
<th>Estimate</th>
<th>Std. Error</th>
<th>( t ) value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>968.5671</td>
<td>17.6221</td>
</tr>
<tr>
<td>PrevTrials</td>
<td>-17.9604</td>
<td>0.8583</td>
</tr>
<tr>
<td>FontSize</td>
<td>12.7588</td>
<td>0.2309</td>
</tr>
</tbody>
</table>
Great! A $t$ value. This will be really helpful for my inferential statistics.

But you also need the degrees of freedom! And degrees of freedom are not exactly defined for mixed effects models. GOT YOU!

But, we can estimate the degrees of freedom.

Curses! Foiled again!
Hypothesis Testing—lmerTest

- Another add-on package, lmerTest, that estimates the d.f. for the t-test
  - Similar to correction for unequal variance

- Tools menu -> Install Packages...
- This time, get lmerTest
Hypothesis Testing—lmerTest

- Once we have `lmerTest` installed, need to load it … remember how?
  - `library(lmerTest)`

- With `lmerTest` loaded, re-run the `lmer()` model, then get its summary
  - Will have p-values

- In the future, no need to run model twice. Can load `lmerTest` from the beginning
  - This was just for demonstration purposes
Hypothesis Testing—lmerTest

| Fixed effects | Estimate | Std. Error | df   | t value | Pr(>|t|)   |
|---------------|----------|------------|------|---------|------------|
| (Intercept)   | 968.5671 | 17.6221    | 138.7503 | 54.96   | <2e-16 *** |
| PrevTrials    | -17.9604 | 0.8583     | 1377.8646 | -20.93  | <2e-16 *** |
| FontSize      | 12.7588  | 0.2309     | 1430.8796 | 55.26   | <2e-16 *** |

ESTIMATED degrees of freedom – note that it’s possible to have non-integer numbers because it’s an estimate

p-value (here, < .0001)
Confidence Intervals

• 95% confidence intervals are:
  • Estimate $\pm (1.96 \times \text{std. error})$

• Try calculating the confidence interval for the **font size** effect
• This is slightly **anticonservative**
  • In other words, with small samples, CI will be too small (elevated risk of Type I error)
  • But OK with even moderately large samples
Confidence Intervals

- Another add-on package: psycholing
- Includes `summaryCI()` function that does this for all fixed effects
Week 3: Fixed Effects

- Installing Packages
- Fixed Effects
  - Introduction to Fixed Effects
  - Running the Model in R
- Hypothesis Testing
  - Model Formulae
  - Interpreting Interactions
  - Model Fitting
  - Fitted Values, Residuals, & Outliers
- Effect Size
  - Unstandardized
  - Standardized
  - Interpretation
  - Overall Variance Explained
Hang on, what if I think that the **font size** and **serial position** will interact?
- Font size effect might get weaker as you get practice with the task

Add an interaction to the model:
- `model2 <- lmer(RT ~ 1 + PrevTrials + FontSize + PrevTrials:FontSize + (1|Subject) + (1|Item), data=Stroop)`
- : means interaction
Model Formulae: Interactions

• A shortcut!
• $1 + \text{PrevTrials} \ast \text{FontSize}$
  • A * means the interaction plus all of the individual effects
  • For factorial experiments (where we use every combination of independent variables), usually what you want
• Try fitting a model3 using * and see if you get the same results as model 2

• Scales up to even more variables: 
  YearsOfStudy*WordFrequency*NounOrVerb
Model Formulae Practice

• What do each of these formulae represent?
  • CollegeGPA ~ 1 + SATScore + HighSchoolGPA
  • PerceivedCausalStrength ~ 1 + PriorBelief + StrengthOfRelation + PriorBelief:StrengthOfRelation
  • DetectionRT ~ 1 + Brightness*Contrast + PreviousTrialRT
Model Formulae Practice

• What do each of these formulae represent?
  • CollegeGPA ~ 1 + SATScore + HighSchoolGPA
    • College GPA predicted by SAT score & high school GPA, no interaction
  • PerceivedCausalStrength ~ 1 + PriorBelief + StrengthOfRelation + PriorBelief:StrengthOfRelation
    • Perceived causal strength predicted by strength of relation, prior belief, and their interaction
  • DetectionRT ~ 1 + Brightness*Contrast + PreviousTrialRT
    • Detection RT predicted by brightness, contrast, & their interaction plus previous trial RT
Model Formulae Practice

• Write the formula for each model:
  • 1) We’re interested in the effects of family SES, prior night’s sleep, and nutrition on math test performance, but we don’t expect them to interact

  • 2) We factorially manipulated sentence type (active or passive) and plausibility in a test of text comprehension accuracy
Model Formulae Practice

• Write the formula for each model:
  • 1) We’re interested in the effects of family SES, prior night’s sleep, and nutrition on math test performance, but we don’t expect them to interact
    • MathPerformance ~ 1 + SES + Sleep + Nutrition
  • 2) We factorially manipulated sentence type (active or passive) and plausibility in a test of text comprehension accuracy
    • ComprehensionAccuracy ~ 1 + SentenceType + Plausibility + SentenceType:Plausibility or
    • ComprehensionAccuracy ~ 1 + SentenceType*S plausibility
Week 3: Fixed Effects

- Installing Packages
- Fixed Effects
  - Introduction to Fixed Effects
  - Running the Model in R
  - Hypothesis Testing
  - Model Formulae
    - Interpreting Interactions
    - Model Fitting
    - Fitted Values, Residuals, & Outliers
- Effect Size
  - Unstandardized
  - Standardized
  - Interpretation
  - Overall Variance Explained
Interpreting Interactions

• Doesn’t look like much of an interaction

\[ y = 954 + -17*\text{PrevTrials} + 13*\text{FontSize} + (-0.02*\text{PrevTrials}*\text{FontSize}) \]

• What would the interaction mean if it existed?

Amplifies the PrevTrials effect (larger number = smaller RT) if large font size

When would this decrease RT the most? (most negative number)

Reduces the FontSize effect (larger number = longer RT) if more previous trials

When prev trials is large

When font size is large
Interpreting Interactions

• Doesn’t look like much of an interaction

<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>954.47234</td>
<td>26.21396</td>
<td>36.41</td>
</tr>
<tr>
<td>PrevTrials</td>
<td>-16.71998</td>
<td>1.90486</td>
<td>-8.78</td>
</tr>
<tr>
<td>FontSize</td>
<td>13.03436</td>
<td>0.44372</td>
<td>29.38</td>
</tr>
<tr>
<td>PrevTrials:FontSize</td>
<td>-0.02414</td>
<td>0.03311</td>
<td>-0.73</td>
</tr>
</tbody>
</table>

• What would the interaction mean if it existed?
  • Negatively-signed interactions (like this one) amplify negatively signed effects and reduce positively signed effects
  • Positively-signed interactions amplify positively signed effects and reduce negatively signed effects
Interpreting Interactions Practice

- Dependent variable: Classroom learning
- Independent variable 1: Intrinsic motivation
  - Learning because you want to learn (intrinsic) vs. to get a good grade (extrinsic)
  - Intrinsic motivation has a + effect on learning
- Independent variable 2: Autonomy language
  - “You can…” (vs. “You must…”)
  - Also has a + effect on learning
- Motivation x autonomy interaction is +
  - Interpretation: Combining intrinsic motivation and autonomy language especially benefits learning
  - “Synergistic” interaction

Vansteenkiste et al., 2004, JPSP
Interpreting Interactions Practice

- Dependent variable: Satisfaction with a consumer purchase
- Number of choices: effect on satisfaction
- "Maximizing" strategy: effect on satisfaction
  - Trying to find the best option vs. "good enough"
- Choices x maximizing strategy is
  - Interpretation: Having lots of choices when you’re a maximizer especially reduces satisfaction
- Also a synergistic interaction

(Carrillat, Ladik, & Legoux, 2011; Marketing Letters)
Interpreting Interactions Practice

- Garden-path sentences:
  - “The horse raced past the barn fell.”
  - = “The horse [that someone] raced past the barn [was the horse that] fell.”
  - “The poster drawn by the illustrator appeared on a magazine cover.”

- Syntactic ambiguity: + effect on reading time (longer reading time)
- Animate (living) subject: No main effect on reading time
- Ambiguity x animacy interaction is +
  - Interpretation: Animate subject not harder by itself, but amplifies the syntactic ambiguity effect

(Trueswell et al., 1994, JML)
**Interpreting Interactions Practice**

- Second language proficiency: + effect on translation accuracy
- Word frequency: + effect on accuracy
- Frequency x proficiency interaction is -
  - Interpretation: Word frequency effect gets smaller if high proficiency
  - (Or: Proficiency matters less when translating high frequency words)
  - “Antagonistic” interaction. Combining the effects reduces or reverses the individual effects.

(e.g., Diependaele, Lemhöfer, Brysbaert, 2012, QJEP)
Interpreting Interactions Practice

- Retrieval practice: + effect on long-term learning
- Low working memory (WM) span: - effect on learning
- Retrieval practice x WM span interaction is + (Agarwal et al., 2016)
  - Interpretation: Retrieval practice is especially beneficial for people with low working memory. (Or: Low WM confers less of a disadvantage if you do retrieval practice.)
Interpreting Interactions Practice

- Affectionate touch: + effect on feeling of relationship security
- Avoidant attachment style: - effect on security
- Touch x avoidant attachment interaction is -
  - Interpretation: Affectionate touch enhances relationship security less for people with an avoidant attachment style

(Jakubiak & Feeney, SPPS, 2016)
Interpreting Interactions Practice

- **Age**: effect on picture memory
  - Older adults have poorer memory
- **Emotional valence**: effect on accuracy
  - Positive pictures are not remembered as well compared to negative pictures

**Age x Valence interaction is +**
- Interpretation: Age declines are smaller for positive pictures
- (Or: Disadvantage of positive pictures is not as strong for older adults)

(e.g., Mather & Carstensen, 2005, *TiCS*)
Interpreting Interactions

- Fixed effect estimates provide a numerical description of the interaction
  - Sufficient to describe the interaction!
  - And, they test the statistical significance
- But, in many cases, looking at a figure of the descriptive statistics will be very helpful for understanding
  - Good to do whenever you’re uncertain
Week 3: Fixed Effects

- Installing Packages
- Fixed Effects
  - Introduction to Fixed Effects
  - Running the Model in R
  - Hypothesis Testing
  - Model Formulae
  - Interpreting Interactions
- Model Fitting
- Fitted Values, Residuals, & Outliers
- Effect Size
  - Unstandardized
  - Standardized
  - Interpretation
- Overall Variance Explained
Model Fitting

- We specified the formula
- How does R know what the right $\gamma$ values are for this model?
Model Fitting

• Solve for $x$:
  • $2(x + 7) = 18$
2(x + 7) = 18

- Two ways you might solve this:
  - **Use algebra**
    1. 2(x+7) = 18
    2. x+7 = 9
    3. x = 2
    4. Guaranteed to give you the right answer

- **Guess and check:**
  1. x = 10? -> 34 = 18 *Way off!*
  2. x = 1? -> 9 = 18 *Closer!*
  3. x = 2? -> 18 = 18 *Got it!*
  4. Might have to check a few numbers
Model Fitting

- Two ways you might solve this:
  - **t-test**: Simple formula you can solve with algebra
    \[ t = \frac{\bar{X}_1 - \bar{X}_2}{s_{X_1X_2} \cdot \sqrt{\frac{2}{n}}} \]
  - **Mixed effects models**: Need to *search* for the best estimates

*ANALYTIC SOLUTION*

*NON-ANALYTIC SOLUTION*
Model Fitting

• In particular, looking for the *model parameters* (results) that have the greatest (log) *likelihood* given the data
  • Maximum likelihood estimation

• Not guessing randomly. Looks for better & better parameters until it *converges* on the solution
  • Like playing “warmer”/“colder”
Model Fitting—Implications

• More complex models take more time to fit
  • `model1 <- lmer(RT ~ 1 + PrevTrials + FontSize + (1|Subject) + (1|Item), data=Stroop, verbose=2)`
  • `verbose=2` shows R’s steps in the search
  • Probably don’t need this; just shows you how it works

• Possible for model to fail to converge on a set of parameters
  • Issue comes up more when you have more complex models (namely, lots of random effects)
  • We’ll talk more in a few weeks about when this might happen & what to do about it
Week 3: Fixed Effects

- Installing Packages
- Fixed Effects
  - Introduction to Fixed Effects
  - Running the Model in R
  - Hypothesis Testing
  - Model Formulae
  - Interpreting Interactions
- Model Fitting
  - Fitted Values, Residuals, & Outliers
- Effect Size
  - Unstandardized
  - Standardized
  - Interpretation
  - Overall Variance Explained
**Predicted Values**

- A model implies a *predicted* value for each observation (“y hat”):
  \[ \hat{y} = 954 + -17 \times \text{PrevTrials} + 13 \times \text{FontSize} \]
- For a trial with 10 previous trials and a font size of 36, what do we predict as the RT?

- See all of the predicted/fitted values:
  - `fitted(model1)`
  - Make them a column in your dataframe:
    - `Stroop$PredictedRT <- fitted(model1)}`
Residuals

- How far off are our individual predictions?
- **Residuals**: Difference between predicted & actual for a specific observation

- “2% or 3% [market share] is what Apple might get.”
  — former Microsoft CEO Steve Ballmer on the iPhone

- **Actual iPhone market share (2014): 42%**
- **Residual: 39 to 40 percentage points**
**Residuals**

- `resid(model1)`
- Residuals are on the same scale as the original DV (e.g., milliseconds or Likert ratings)
  - `abs(scale(resid(model1))`
    - z-scores them so they’re in *number of standard deviations*
- Can use this to identify & remove outliers
  - `Stroop.OutliersRemoved <- Stroop[abs(scale(resid(model1))) <= 3, ]`
  - Outliers *after* accounting for all of the variables of interest, subjects, and items
    - Long RT might not be an outlier if slowest subject on slowest item
- How many data points did we lose?
  - `nrow(naming) - nrow(naming.OutliersRemoved)`
**How Should Outliers Change Interpretation?**

- Effect reliable with and without outliers?
  - Hooray!

- Effect only seen if outliers removed?
  - Effect characterizes *most* of the data, but a few exceptions

- Effect only seen with outliers included?
  - Suggests it’s driven by a few observations

- No effect either way?
  - Weep softly at your desk
Week 3: Fixed Effects

- Installing Packages
- Fixed Effects
  - Introduction to Fixed Effects
  - Running the Model in R
  - Hypothesis Testing
  - Model Formulae
  - Interpreting Interactions
  - Model Fitting
  - Fitted Values, Residuals, & Outliers
- Effect Size
  - Unstandardized
  - Standardized
  - Interpretation
  - Overall Variance Explained
Effect Size

• Remember that $t$ statistics and p-values tell us about whether there’s an effect in the population
  • Is the effect statistically reliable?

• A separate question is how big the effect is
  • Effect size
Bigfoot: Little evidence he exists, but he'd be large if he did exist.

Sample 1: [-.20, 1.80]  
LARGE EFFECT SIZE, LOW RELIABILITY

Sample 2: [.15, .35]  
SMALL EFFECT SIZE, HIGH RELIABILITY

Pygmy hippo: We know it exists and it's small.
• Is bacon really this bad for you??

October 26, 2015
Is bacon really this bad for you??

True that we have as much evidence that bacon causes cancer as smoking causes cancer!

Same level of statistical reliability
• Is bacon really this bad for you??
• True that we have as much evidence that bacon causes cancer as smoking causes cancer!
  • Same level of statistical reliability
• But, effect size is much smaller for bacon
Effect Size

• Our model results tell us both

Parameter estimate tells us about effect size

t statistic and p-value tell us about statistical reliability

Fixed effects:

|          | Estimate | Std. Error | df    | t value | Pr(>|t|) |
|----------|----------|------------|-------|---------|----------|
| (Intercept) | 968.5671 | 17.6221    | 138.7000 | 54.96   | <2e-16 *** |
| PrevTrials  | -17.9604 | 0.8583     | 1365.9000 | -20.93  | <2e-16 *** |
| FontSize    | 12.7588  | 0.2309     | 1418.7000 | 55.26   | <2e-16 *** |
**Effect Size: Parameter Estimate**

- Simplest measure: Parameter estimates
  - Effect of 1-unit change in predictor on outcome variable
  - “On average, RT decreased by 18 ms for each additional trial of experience”
  - “Each minute of exercise increases life expectancy by about 7 minutes.” (Moore et al., 2012, *PLOS ONE*)
  - “People with a college diploma earn around $24,000 more per year.” (Bureau of Labor Statistics, 2018)
  - Concrete! Good for “real-world” outcomes

| Fixed effects          | Estimate | Std. Error | df  | t value | Pr(>|t|) |
|------------------------|----------|------------|-----|---------|----------|
| (Intercept)            | 968.5671 | 17.6221    | 138.7503 | 54.96   | <2e-16 *** |
| PrevTrials             | -17.9604 | 0.8583     | 1377.8646 | -20.93  | <2e-16 *** |
| FontSize               | 12.7588  | 0.2309     | 1430.8796 | 55.26   | <2e-16 *** |
Week 3: Fixed Effects

- Installing Packages
- Fixed Effects
  - Introduction to Fixed Effects
  - Running the Model in R
  - Hypothesis Testing
  - Model Formulae
  - Interpreting Interactions
  - Model Fitting
  - Fitted Values, Residuals, & Outliers
- Effect Size
  - Unstandardized
  - Standardized
  - Interpretation
  - Overall Variance Explained
Effect Size: Standardization

• Which is the bigger effect?
  • 1 minute of exercise = 7 minutes of life expectancy
  • Smoking 1 pack of cigarettes = -11 minutes of life expectancy (Shaw, Mitchell, & Dorling, 2000, BMJ)

• Problem: These are measured in different units
  • Minutes of exercise vs. packs of cigarettes

• Convert to **z-scores**: # of standard deviations from the mean
  • This scale applies to anything!
  • **Standardized** scores
**Effect Size: Standardization**

- `scale()` puts things in terms of z-scores
- New z-scored version of FontSize:
  - `Stroop$FontSize.z <- scale(Stroop$FontSize)[,1]`
  - # of standard deviations above/below mean font size)

- Do the same for RT and FontSize
- Then use them in a new model
Effect Size: Standardization

• My results:

But, **effect size** is now estimated differently

Notice the $t$ statistics for our critical effects have *not* changed ... no change in **statistical reliability**

<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>0.387e-17</td>
<td>2.055e-02</td>
<td>0.00</td>
</tr>
<tr>
<td>FontSize.z</td>
<td>7.890e-01</td>
<td>1.428e-02</td>
<td>55.26</td>
</tr>
<tr>
<td>PrevTrials.z</td>
<td>-2.962e-01</td>
<td>1.415e-02</td>
<td>-20.93</td>
</tr>
</tbody>
</table>
Interlude: Scientific Notation

• OK, but what’s all of this e nonsense!?
• Scientific notation
  • $7.890e^{-01}$ is $7.89 \times 10^{-1} = 0.789$
  • $e^{-xx}$ = Move the decimal place $xx$ numbers to the left (smaller number)
  • $e^{+xx}$ = Move the decimal place $xx$ numbers to the right (larger number)
Interlude: Scientific Notation

• Scientific notation is a good way to write really small numbers, like $6.387 \times 10^{-17}$
• That’s $6.387 \times 10^{-17}$
• Intercept is practically zero … when at average font size & average serial position (z-scores of 0), RT is also average (z-score of 0)
• True by definition when using z-scores
Interlude: Scientific Notation

- Scientific notation is a good way to write really small numbers, like $6.387\times10^{-17}$
- When at least one number in your results needs scientific notation, R uses it throughout
- Can just copy & paste these into R prompt to translate them:

```r
> -2.962e-01
[1] -0.2962
```

Fixed effects:

<table>
<thead>
<tr>
<th>Estimate</th>
<th>Std. Error</th>
<th>t value</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>$6.387e-17$</td>
<td>2.055e-02</td>
<td>0.00</td>
</tr>
<tr>
<td>FontSize.z</td>
<td>7.890e-01</td>
<td>1.428e-02</td>
<td>55.26</td>
</tr>
<tr>
<td>PrevTrials.z</td>
<td>-2.962e-01</td>
<td>1.415e-02</td>
<td>-20.93</td>
</tr>
</tbody>
</table>
Effect Size: Standardization

- Which of our two critical effects has the effect size of larger magnitude? (disregarding the direction)
  - 1 standard deviation change in font size = Increase of .789 standard deviations in RTs
  - 1 standard deviation change in serial position = Decrease of .296 standard deviations in RTs

Fixed effects:

|                | Estimate | Std. Error | t value | Pr(>|t|) |
|----------------|----------|------------|---------|----------|
| (Intercept)    | 6.387e-17| 2.055e-02  | 0.00    |          |
| FontSize.z     | 7.890e-01| 1.428e-02  | 55.26   | < 2e-16  |
| PrevTrials.z   | -2.962e-01| 1.415e-02 | -20.93  | < 2e-16  |
Effect Size: Standardization

Which of our two critical effects has the **effect size** of larger magnitude? (disregarding the direction)

- 1 standard deviation change in font size = Increase of .789 standard deviations in RTs
- 1 standard deviation change in serial position = Decrease of .296 standard deviations in RTs

![Fixed effects table]

<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>6.387e-17</td>
<td>2.055e-02</td>
<td>0.00</td>
</tr>
<tr>
<td>FontSize.z</td>
<td>7.890e-01</td>
<td>1.428e-02</td>
<td>55.26</td>
</tr>
<tr>
<td>PrevTrials.z</td>
<td>-2.962e-01</td>
<td>1.415e-02</td>
<td>-20.93</td>
</tr>
</tbody>
</table>
**Effect Size: Standardization**

- But, standardized effects make our effect sizes somewhat more reliant on our data.
- Effect of 1 std dev of cigarette smoking on life expectancy depends on what that std. dev is.
  - Varies a lot from country to country!
  - Might get different standardized effects even if unstandardized is the same.
Week 3: Fixed Effects

- Installing Packages
- Fixed Effects
  - Introduction to Fixed Effects
  - Running the Model in R
- Hypothesis Testing
- Model Formulae
- Interpreting Interactions
- Model Fitting
- Fitted Values, Residuals, & Outliers
- Effect Size
  - Unstandardized
  - Standardized
  - Interpretation
  - Overall Variance Explained
Effect Size: Interpretation

• Generic heuristic for standardized effect sizes
  • “Small” ≈ .25
  • “Medium” ≈ .50
  • “Large” ≈ .80

• But, take these with several grains of salt
  • Cohen (1988) just made them up
  • Not in context of particular domain
Effect Size: Interpretation

- Consider in context of other effect sizes in this domain:

  Our effect: .20  Other effect 1: .30  Other effect 2: .40
  Other effect 1: .10  Other effect 2: .15  Our effect: .20

- vs:

- For interventions: Consider cost, difficulty of implementation, etc.
  - Aspirin’s effect in reducing heart attacks: $d \approx .06$, but cheap!
Effect Size: Interpretation

- For theoretically guided research, compare to predictions of competing theories

- The lag effect in memory:

  - Is this about intervening items or time?
**Effect Size: Interpretation**

- Is lag effect about intervening **items** or **time**?

  - Intervening items hypothesis predicts A > B
  - Time hypothesis predicts B > A
  - Goal here is to use direction of the effect to adjudicate between competing hypotheses
  - *Not* whether the lag effect is “small” or “large”
Week 3: Fixed Effects

- Installing Packages
- Fixed Effects
  - Introduction to Fixed Effects
  - Running the Model in R
  - Hypothesis Testing
  - Model Formulae
  - Interpreting Interactions
  - Model Fitting
  - Fitted Values, Residuals, & Outliers
- Effect Size
  - Unstandardized
  - Standardized
  - Interpretation
- Overall Variance Explained
Overall Variance Explained

- How well do predicted values match up with what actually happened?
  - How well did we explain the outcomes?
- $R^2$:
  \[
  \text{cor(fitted(model1), Stroop$RT)^2}
  \]
  - But, this includes what’s predicted on basis of subjects/items
  - Compare to the $R^2$ of a model with just the subjects & items
Week 3: Fixed Effects

- Installing Packages
- Fixed Effects
  - Introduction to Fixed Effects
  - Running the Model in R
  - Hypothesis Testing
  - Model Formulae
  - Interpreting Interactions
  - Model Fitting
  - Fitted Values, Residuals, & Outliers
- Effect Size
  - Unstandardized
  - Standardized
  - Interpretation
  - Overall Variance Explained
Conclusion

• Fixed effects are the variables of interest
  • Estimated with maximum likelihood estimation
  • Characterize relation of predictors to outcome
  • Defined by model formula
  • Can test their contribution to the model
    • $z$-test, $t$-test with estimated degrees of freedom
  • Residuals can help detect outliers
  • Fixed effect estimates tell us about effect size

• Next week: Model comparison & random effects