Course Business

- Midterm assignment now posted on CourseWeb
  - Due October 21st – 3 weeks from today

- Recognize some aspects of last week’s class were especially challenging
  - We will revisit categorical DVs in the second half of the course

- Upcoming classes:
  - 10/7: Categorical predictors with >2 categories, pros & cons of mixed effects models
  - 10/14: No class
  - 10/21: Discuss papers
Week 6: Coding Predictors I

- Centering Continuous Variables
- Introduction to Contrast Coding
- Dummy Coding
  - What it Means
  - How to Change Codes
  - Interactions
- Effects Coding
- Main Effects vs Simple Main Effects
- Unbalanced Factors
The Big Picture

- Choices about how to code variables (esp. categorical ones)
  - Allow us to answer *different questions* about the data

- In most cases, multiple statistically valid ways to code
  - But, important that we actually perform the test that corresponds to *what we say we want to know*
http://www.scottfraundorf.com/R/aphasia.csv

- Task: Decide whether a picture matches a sentence; measure RT

“The dog was chased by the man.”
Task: Decide whether a picture matches a sentence; measure RT

Between-subjects variables (within-items):
- subjectType: Aphasia or Control

Between-items variables (within-subjects):
- sentenceLength (in words)
- sentenceType: Active or Passive
  - Active (more common): “Man bites dog.”
  - Passive: “The dog was bitten by the man.”
Interpreting Intercepts

- Let's do a simple model of the effect of sentence length on RT
- `lengthModel.Maximal <- lmer(RT ~ 1 + SentenceLength +
  (1 + SentenceLength|Subject) +
  (1|Item), data = aphasia)`
**Interpreting Intercepts**

- Let’s do a simple model of the effect of sentence length on RT

\[
\text{lengthModel.Maximal} \leftarrow \text{lmer}(\text{RT} \sim 1 + \text{SentenceLength} + (1 + \text{SentenceLength}|\text{Subject}) + \text{ITEM RANDOM EFFECTS})
\]

- data = aphasia

- Sentence length is manipulated within subjects (each subject sees several different sentence lengths)
  - Possible to calculate each subject’s personal SentenceLength effect (slope)
  - Include random slope
Interpreting Intercepts

- Let’s do a simple model of the effect of sentence length on RT
- `lengthModel.Maximal <- lmer(RT ~ 1 + SentenceLength +
  (1 + SentenceLength|Subject) +
  (1|Item),
data = aphasia)`

- Sentence length is manipulated between items (each sentence has only one length)
  - Not possible to calculate a SentenceLength effect (slope) using just one item
  - Don’t include random slope
Interpreting Intercepts

Let’s do a simple model of the effect of sentence length on RT

\[
\text{lengthModel.Maximal} \leftarrow \text{lmer}(\text{RT} \sim 1 + \text{SentenceLength} + (1 + \text{SentenceLength}|\text{Subject}) + (1|\text{Item}), \\
data = \text{aphasia})
\]

The statistician's cheer:
- When I say “within,” you say “random slope!”
Interpreting Intercepts

• Let’s do a simple model of the effect of sentence length on RT

• Results:

\[ y = 1215 + 88 \times \text{SentenceLength} \]

• Intercept: RT is 1215 ms when sentence length is 0

• Sentence length effect: +88 ms for each word

• But, sentence length 0 is impossible. Odd to talk about.
Let’s change the model so that 0 means something
Mean Centering

- Mean sentence length is **10.00**

- Imagine we subtracted **this mean** length from **each sentence length**

<table>
<thead>
<tr>
<th>ORIGINAL</th>
<th>MEAN SUBTRACTED</th>
</tr>
</thead>
<tbody>
<tr>
<td>Traffic jam: 7</td>
<td>-3</td>
</tr>
<tr>
<td>Chess club: 10</td>
<td>0</td>
</tr>
<tr>
<td>Panther: 11</td>
<td>1</td>
</tr>
</tbody>
</table>

- **New zero** represents **mean** length
- “Mean centering”
Mean Centering

- *New zero* represents *mean* length
- “*Mean centering*”
Centering—How to Do It

• **First**, create a new variable:
  • `aphasia$SentenceLength.cen = scale(aphasia$SentenceLength, center=TRUE, scale=FALSE)[,1]`

• **Then**, use the new variable in your model
  • `lengthModel.cen.Maximal <- lmer(RT ~ 1 + SentenceLength.cen + (1 + SentenceLength.cen|Subject) + (1|Item), data = aphasia)`
Centering—Results

• Old model:

<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>1215.43</td>
<td>373.96</td>
<td>3.250</td>
</tr>
<tr>
<td>SentenceLength</td>
<td>87.86</td>
<td>36.82</td>
<td>2.386</td>
</tr>
</tbody>
</table>

\[ y = 2094 + 88 \times \text{SentenceLength} \]

• New model:

<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>2094.04</td>
<td>78.24</td>
<td>26.765</td>
</tr>
<tr>
<td>SentenceLength.cen</td>
<td>87.86</td>
<td>36.82</td>
<td>2.386</td>
</tr>
</tbody>
</table>

• Intercept: RT is 2094 ms at mean sentence length
• Sentence length effect: +88 ms for each add’l word
Centering—Results

• Old model:

<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>1215.43</td>
<td>373.96</td>
<td>3.250</td>
</tr>
<tr>
<td>SentenceLength</td>
<td>87.86</td>
<td>36.82</td>
<td>2.386</td>
</tr>
</tbody>
</table>

Correlation of Fixed Effects:
- (Intr) 0.978
- SentncLnth -0.978

Correlation of sentence length effect with intercept is now almost 0. Indicates that we centered correctly.

• New model:

<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>2094.04</td>
<td>78.24</td>
<td>26.765</td>
</tr>
<tr>
<td>SentenceLength.cen</td>
<td>87.86</td>
<td>36.82</td>
<td>2.386</td>
</tr>
</tbody>
</table>

Correlation of Fixed Effects:
- (Intr) 0.033
- SentncLnth 0.033

\[ y = 2094 + 88 \times \text{SentenceLength} \]

• Intercept: RT is 2094 ms at mean sentence length
• Sentence length effect: +88 ms for each add’l word
Centering—Results

- Intercept: RT is **2094** ms at mean sentence length
- Sentence length effect: +88 ms for each add’l word
Centering—Results

- Both regression equations apply only to plausible sentence lengths
  - With raw sentence length, can’t have a sentence length less than 0 (no such thing as a negative # of words!)
  - With centered sentence length, can’t have a sentence length less than -10 (0 minus the mean of 10)
Which Do You Like Better?

**UNCENTERED**
- Good if zero is *meaningful*
- Years of study abroad, number of previous trials, number of missed classes…

**CENTERED**
- Good if zero is *not meaningful* or *not observed*
- Reduced correlation w/ intercept also helps with convergence (esp. in binomial models)
Other Alternatives

• It would also be possible to make 0 correspond to some other sensible/useful value
  • e.g., 0 could be the *shortest* sentence length in our set of items
  • `aphasia$SentenceLength2 <- aphasia$SentenceLength - min(aphasia$SentenceLength)`
Week 6: Coding Predictors I

- Centering Continuous Variables
- Introduction to Contrast Coding
- Dummy Coding
  - What it Means
  - How to Change Codes
  - Interactions
- Effects Coding
- Main Effects vs Simple Main Effects
- Unbalanced Factors
**Terminology**

- **Factor:** A categorical variable
  - Variables where we get counts in our R summary
  - `as.factor()` makes things categorical if they aren’t already

- **Levels:** The individual categories within a factor
  - “Active” versus “Passive”
  - “Aphasia” versus “Healthy control”
Introduction to Contrast Coding

• So far, we’ve been writing regression equations with numbers

\[ RT = \text{Intercept} + Y_{100} \]

• But what about active vs passive sentence?

\[ RT = \text{Intercept} + Y_{100} \]
Introduction to Contrast Coding

• But what about active vs passive sentence?

\[
RT = \text{Intercept} + \gamma_{100}
\]

• R’s “secret decoder wheel” assigns numerical coding schemes:
  • Variable with 2 categories (this week): Only one comparison needed
  • Variables with more categories: Multiple contrasts

Active sentence: 0
Passive sentence: 1
Introduction to Contrast Coding

• But what about active vs passive sentence?

R’s “secret decoder wheel” assigns numerical coding schemes

• See the current codes:
  • `contrasts(aphasia$SentenceType)`

\[
RT = \text{Intercept} + Y_{100}
\]
Week 6: Coding Predictors I

- Centering Continuous Variables
- Introduction to Contrast Coding
- Dummy Coding
  - What it Means
  - How to Change Codes
  - Interactions
- Effects Coding
- Main Effects vs Simple Main Effects
- Unbalanced Factors
Treatment Coding (Dummy Coding)

- R’s default system
  - One baseline/reference **level** (category is coded as 0)
  - The other (the **treatment**) is coded as 1
- Default ordering is alphabetical: First level is 0, second is 1
  - We’ll see how to change this soon

- `contrasts(aphasia$SentenceType)`
  - Active coded as 0
  - Passive coded as 1
Treatment Coding (Dummy Coding)

• Let’s do a model that just examines the effect of sentence type in this task:
• `SentenceTypeModel <- lmer(RT ~ 1 + SentenceType + SUBJECT RANDOM EFFECTS
  ITEM RANDOM EFFECTS
data = aphasia)`
Let’s do a model that just examines the effect of sentence type in this task:

\[
\text{SentenceTypeModel} \leftarrow \text{lmer}(\text{RT} \sim 1 + \text{SentenceType} + (1 + \text{SentenceType}|\text{Subject}) + \text{ITEM RANDOM EFFECTS}, \text{data} = \text{aphasia})
\]
Treatment Coding (Dummy Coding)

• Let’s do a model that just examines the effect of sentence type in this task:
• `SentenceTypeModel <- lmer(RT ~ 1 + SentenceType + (1 + SentenceType|Subject) + (1|Item), data = aphasia)`
Treatment Coding (Dummy Coding)

- Let’s think about what the model looks like for each of our two conditions:

\[
RT = \gamma_{000} + \gamma_{100} \times \text{SentenceType}
\]

Active Sentences

Passive Sentences
Treatment Coding (Dummy Coding)

- Let’s think about what the model looks like for each of our two conditions:

Active Sentences: \[ RT = \gamma_{000} + \gamma_{100} \]

Passive Sentences: \[ RT = \gamma_{000} + \gamma_{100} \]
Treatment Coding (Dummy Coding)

Let’s think about what the model looks like for each of our two conditions:

- **Active Sentences**
  \[ RT = \gamma_{000} + \gamma_{100} \]

- **Passive Sentences**
  \[ RT = \gamma_{000} + \gamma_{100} \]
Treatment Coding (Dummy Coding)

• Let’s think about what the model looks like for each of our two conditions:

<table>
<thead>
<tr>
<th>Condition</th>
<th>Model Equation</th>
<th>Intercept</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active Sentences</td>
<td>( RT = \gamma_{000} )</td>
<td>Intercept is just the mean RT for active sentences</td>
</tr>
<tr>
<td>Passive Sentences</td>
<td>( RT = \gamma_{000} + \gamma_{100} )</td>
<td>1</td>
</tr>
</tbody>
</table>
**Treatment Coding (Dummy Coding)**

- Let’s think about what the model looks like for each of our two conditions:

  - **Active Sentences**
    \[ RT = \gamma_{000} \]
    
    Intercept is just the mean RT for active sentences

  - **Passive Sentences**
    \[ RT = \gamma_{000} + \gamma_{100} \]
    
    Sentence Type effect is the difference in RT between passive & active sentences

What is the difference between the equations for the two sentence types?
Treatment Coding Results

<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>1758.27</td>
<td>66.02</td>
<td>26.633</td>
</tr>
<tr>
<td>SentenceTypePassive</td>
<td>671.54</td>
<td>78.49</td>
<td>8.556</td>
</tr>
</tbody>
</table>

Correlation of Fixed Effects:

| (Intr)                        |           |             |         |
| SntncTypPss                   | -0.521    |             |         |

- Treatment coding makes one level the baseline and compares everything to that

Intercept: RT for active sentences is 1758 ms
SentenceType: RT difference between conditions is 672 ms
Week 6: Coding Predictors I

- Centering Continuous Variables
- Introduction to Contrast Coding
- Dummy Coding
  - What it Means
    - How to Change Codes
    - Interactions
- Effects Coding
- Main Effects vs Simple Main Effects
- Unbalanced Factors
Changing Codes

- We should think about adding SubjectType to the model. Let's check the codes:
  - `contrasts(aphasia$SubjectType)`

```
Control
Aphasia  0
Control  1
```

- But, control is really the baseline category here
- Assign new codes by using `<-`:
  - `contrasts(aphasia$SubjectType) <- c(1,0)`
- New codes are in the order you see above & with `summary()`
Changing Codes

• Need to set codes **before** you run the model!
Week 6: Coding Predictors I

- Centering Continuous Variables
- Introduction to Contrast Coding
- Dummy Coding
  - What it Means
  - How to Change Codes
  - Interactions
- Effects Coding
- Main Effects vs Simple Main Effects
- Unbalanced Factors
Now, we’d like SentenceType and SubjectType to interact:

```r
Model.Maximal <- lmer(RT ~ 1 + SentenceType * SubjectType +
                       SUBJ ect RANDOM EFFECTS
                       ITEM RANDOM EFFECTS

data = aphasia)
```
Now, we’d like SentenceType and SubjectType to interact:

```r
Model.Maximal <- lmer(RT ~ 1 + SentenceType * SubjectType + (1 + SentenceType|Subject) + (1 + SubjectType|Item), data = aphasia)
```

ITEM RANDOM EFFECTS
Now, we’d like SentenceType and SubjectType to interact:

```r
Model.Maximal <- lmer(RT ~ 1 + SentenceType * SubjectType +
(1 + SentenceType|Subject) +
(1 + SubjectType|Item),
data = aphasia)
```
Treatment Coding: Two Variables

- Our design now has four cells:

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active, Control Subj.</td>
<td>$RT = \gamma_{000} + \gamma_{100}SentenceType + \gamma_{200}SubjectType + \gamma_{1200}SentenceTypeSubjectType$</td>
</tr>
<tr>
<td>Passive, Control Subj.</td>
<td>$RT = \gamma_{000} + \gamma_{100}SentenceType + \gamma_{200}SubjectType + \gamma_{1200}SentenceTypeSubjectType$</td>
</tr>
<tr>
<td>Active, Aphasics</td>
<td>$RT = \gamma_{000} + \gamma_{100}SentenceType + \gamma_{200}SubjectType + \gamma_{1200}SentenceTypeSubjectType$</td>
</tr>
<tr>
<td>Passive, Aphasics</td>
<td>$RT = \gamma_{000} + \gamma_{100}SentenceType + \gamma_{200}SubjectType + \gamma_{1200}SentenceTypeSubjectType$</td>
</tr>
</tbody>
</table>
**Treatment Coding: Two Variables**

- Our design now has four cells:

<table>
<thead>
<tr>
<th>Treatment</th>
<th>RT Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active, Control Subj.</td>
<td>$RT = \gamma_{000} + \gamma_{100} + \gamma_{200} + \gamma_{1200} \cdot 0$</td>
</tr>
<tr>
<td>Passive, Control Subj.</td>
<td>$RT = \gamma_{000} + \gamma_{100} + \gamma_{200} + \gamma_{1200} \cdot 0$</td>
</tr>
<tr>
<td>Active, Aphasics</td>
<td>$RT = \gamma_{000} + \gamma_{100} + \gamma_{200} + \gamma_{1200} \cdot 0$</td>
</tr>
<tr>
<td>Passive, Aphasics</td>
<td>$RT = \gamma_{000} + \gamma_{100} + \gamma_{200} + \gamma_{1200} \cdot 0$</td>
</tr>
</tbody>
</table>
Treatment Coding: Two Variables

- Our design now has four cells:

Active, Control Subj.

\[ RT = \gamma_{000} \]

Passive, Control Subj.

\[ RT = \gamma_{000} + \gamma_{100} + \gamma_{1200} + \gamma_{200} \]

Active, Aphasics

\[ RT = \gamma_{000} + \gamma_{100} \text{SentenceType} + \gamma_{200} \text{SubjectType} + \gamma_{1200} \text{SentenceTypeSubjectType} \]

Passive, Aphasics

\[ RT = \gamma_{000} + \gamma_{100} \text{SentenceType} + \gamma_{200} \text{SubjectType} + \gamma_{1200} \text{SentenceTypeSubjectType} \]

Intercept is the RT when all variables at their baseline: active sentence type, healthy control subject
Treatment Coding: Two Variables

- Our design now has four cells:

  - **Active, Control Subj.**
    \[ RT = \gamma_{000} \]
    \[ \text{Intercept is the RT when all variables at their baseline: active sentence type, healthy control subject} \]

  - **Passive, Control Subj.**
    \[ RT = \gamma_{000} + \gamma_{100} \]
    \[ \text{SentenceType: Passive vs active difference for baseline healthy controls} \]

  - **Active, Aphasics**
    \[ RT = \gamma_{000} + \gamma_{100} + \gamma_{200} \]
    \[ \text{SentenceType + SubjectType + SentenceTypeSubjectType} \]

  - **Passive, Aphasics**
    \[ RT = \gamma_{000} + \gamma_{100} + \gamma_{200} + \gamma_{1200} \]
Treatment Coding: Two Variables

- Our design now has four cells:

\[
RT = \gamma_{000}
\]

- Intercept is the RT when all variables at their baseline: active sentence type, healthy control subject

\[
RT = \gamma_{000} + \gamma_{100}
\]

- SentenceType: Passive vs active difference for baseline healthy controls

\[
RT = \gamma_{000} + \gamma_{200}
\]

- SubjectType: Aphasia vs control difference for baseline active sentences

\[
RT = \gamma_{000} + \gamma_{100} + \gamma_{200} + \gamma_{1200}
\]

- SentenceType + SubjectType + SentenceTypeSubjectType
Treatment Coding: Two Variables

- Our design now has four cells:

  - **Active, Control Subj.**
    - \( RT = \gamma_{000} \)

  - **Passive, Control Subj.**
    - \( RT = \gamma_{000} + \gamma_{100} \)

  - **Active, Aphasics**
    - \( RT = \gamma_{000} + \gamma_{200} \)

  - **Passive, Aphasics**
    - \( RT = \gamma_{000} + \gamma_{100} + \gamma_{200} + 1 \)

  Intercept is the RT when all variables at their baseline: active sentence type, healthy control subject.

  SentenceType: Passive vs active difference for baseline healthy controls.

  SubjectType: Aphasia vs control difference for baseline active sentences.

  If no special effect of passive sentence and aphasia, we’d just have these two effects.
Treatment Coding: Two Variables

- Our design now has four cells:

<table>
<thead>
<tr>
<th>Treatment</th>
<th>RT Expression</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active, Control Subj.</td>
<td>$RT = \gamma_{000}$</td>
<td>Intercept is the RT when all variables at their baseline: active sentence type, healthy control subject</td>
</tr>
<tr>
<td>Passive, Control Subj.</td>
<td>$RT = \gamma_{000} + \gamma_{100}$</td>
<td>Sentence Type: Passive vs active difference for baseline healthy controls</td>
</tr>
<tr>
<td>Active, Aphasics</td>
<td>$RT = \gamma_{000} + \gamma_{200}$</td>
<td>Subject Type: Aphasia vs control difference for baseline active sentences</td>
</tr>
<tr>
<td>Passive, Aphasics</td>
<td>$RT = \gamma_{000} + \gamma_{100} + \gamma_{200}$</td>
<td></td>
</tr>
</tbody>
</table>

If no special effect of passive sentence and aphasia, we’d just have these two effects.
## Treatment Coding: Two Variables

- Our design now has four cells:

<table>
<thead>
<tr>
<th>Active, Control Subj.</th>
<th>Active, Aphasics</th>
<th>Passive, Control Subj.</th>
<th>Passive, Aphasics</th>
</tr>
</thead>
<tbody>
<tr>
<td>$RT = \gamma_{000}$</td>
<td>$RT = \gamma_{000} + \gamma_{200}$</td>
<td>$RT = \gamma_{000} + \gamma_{100}$</td>
<td>$RT = \gamma_{000} + \gamma_{100} + \gamma_{200} + \gamma_{1200}$</td>
</tr>
</tbody>
</table>

Intercept is the RT when all variables at their baseline: active sentence type, healthy control subject

SentenceType: Passive vs active difference for baseline healthy controls

SubjectType: Aphasia vs control difference for baseline active sentences

Interaction: Special effect of aphasia and passive sentence
Treatment Coding: Two Variables

- Our design now has four cells:

  - **Active, Control Subj.**
    
    \[ RT = \gamma_{000} \]
    
    - Intercept is the RT when all variables at their baseline: active sentence type, healthy control subject

  - **Passive, Control Subj.**
    
    \[ RT = \gamma_{000} + \gamma_{100} \]
    
    - SentenceType: Passive vs active difference for baseline healthy controls

  - **Active, Aphasics**
    
    \[ RT = \gamma_{000} + \gamma_{200} \]
    
    - SubjectType: Aphasia vs control difference for baseline active sentences

  - **Passive, Aphasics**
    
    \[ RT = \gamma_{000} + \gamma_{100} + \gamma_{200} + \gamma_{1200} \]
    
    - Interaction: Special effect of aphasia and passive sentence
Treatment Coding: Model Results

Intercept: RT for healthy controls, active voice sentences

Significant RT difference for passive sentences (among healthy controls)

Not a significant RT difference for aphasics (among active sentences)

Significant special effect of aphasia + passive sentence
Treatment Coding: Model Results

Even though the SubjectType effect is not significant here, we would *not* want to remove it from the model. It doesn’t make sense to include the interaction without the lower-order terms—the interaction is defined by what’s different from the two simple effects alone.

### Random effects:

<table>
<thead>
<tr>
<th>Groups</th>
<th>Name</th>
<th>Variance</th>
<th>Std.Dev.</th>
<th>Corr</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item</td>
<td>(Intercept)</td>
<td>47858.30</td>
<td>218.765</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SubjectType1</td>
<td>55.88</td>
<td>7.475</td>
<td>-1.00</td>
</tr>
<tr>
<td>Subject</td>
<td>(Intercept)</td>
<td>43287.42</td>
<td>208.056</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SentenceTypePassive</td>
<td>1443.33</td>
<td>37.991</td>
<td>0.27</td>
</tr>
<tr>
<td>Residual</td>
<td></td>
<td>6850.06</td>
<td>82.765</td>
<td></td>
</tr>
</tbody>
</table>

Number of obs: 960, groups: Item, 32; Subject, 30

### 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>1716.01</td>
<td>76.85</td>
<td>22.330</td>
</tr>
<tr>
<td>SentenceTypePassive</td>
<td>577.17</td>
<td>78.33</td>
<td>7.368</td>
</tr>
<tr>
<td>SubjectType1</td>
<td>84.52</td>
<td>76.37</td>
<td>1.107</td>
</tr>
<tr>
<td>SentenceTypePassive:SubjectType1</td>
<td>188.75</td>
<td>17.71</td>
<td>10.659</td>
</tr>
</tbody>
</table>

Correlation of Fixed Effects:
- (Intr) SntnTP SbjcT1
- SntncTypPss -0.478
- SubjectTyp1 -0.514 -0.002
- SntncTP:ST1 -0.008 -0.258 0.165
Week 6: Coding Predictors I

- Centering Continuous Variables
- Introduction to Contrast Coding
- Dummy Coding
  - What it Means
  - How to Change Codes
  - Interactions
- Effects Coding
- Main Effects vs Simple Main Effects
- Unbalanced Factors
Effects Coding (Sum Coding)

- So far, the intercept at 0 has referred to a particular baseline level
- Remember centering?
  - When we centered, we made the intercept at 0 correspond to the overall mean
Effects Coding / Sum Coding

• We can apply centering to a factor, too

• SentenceType has:
  • 480 “Active” observations (currently 0)
  • 480 “Passive”s (currently 1)

• Mean of 0.5

• Subtracting the mean from each code gives us a new set of codes

• Effects coding (a/k/a sum coding): -0.5, 0.5
**Effects Coding (Sum Coding)**

- Apply effects coding (-0.5, 0.5) to our two sentence types:

  - Active Sentences: $RT = \gamma_{000} + \gamma_{100}$
  - Passive Sentences: $RT = \gamma_{000} + \gamma_{100}$
Effects Coding (Sum Coding)

• Apply effects coding (-0.5, 0.5) to our two sentence types:

Active Sentences

$$RT = \gamma_{000} + \gamma_{100}$$

Passive Sentences

$$RT = \gamma_{000} + \gamma_{100}$$

Improve subtracting the equations.

The difference between the equations for the two conditions is equal to what?
Effects Coding (Sum Coding)

- Apply effects coding (-0.5, 0.5) to our two sentence types:

Active Sentences

\[ \text{RT} = \gamma_{000} + \gamma_{100} \]

Passive Sentences

\[ \text{RT} = \gamma_{000} + \gamma_{100} \]

The equations differ by 1 \( \gamma_{100} \)

SentenceType effect is (still) the difference between conditions

Intercept is always present. It’s now the mean RT across all conditions.
Effects Coding (Sum Coding)

- Let’s apply effects coding to our aphasia data
- Old codes:
  - Active 0
  - Passive 1

- New codes:
  - $\text{contrasts(aphasia}\$\text{SentenceType)} = c(-0.5,0.5)$
  - $\text{contrasts(aphasia}\$\text{SubjectType)} = c(0.5,-0.5)$

- Rerun the model:
  - `EffectsCoding.Maximal <- lmer(RT ~ 1 + SentenceType * SubjectType + (1 + SentenceType|Subject) + (1 + SubjectType|Item), data = aphasia)`
**Effects Coding: Model Results**

**Random effects:**
- **Groups**
  - Item (Intercept): Variance 46236.97, Std.Dev. 215.028
  - SubjectType1: Variance 55.88, Std.Dev. 7.475
- **Subject**
  - Item (Intercept): Variance 45780.92, Std.Dev. 213.965
  - SentenceType1: Variance 1443.33, Std.Dev. 37.991
- **Residual:** Variance 6850.06, Std.Dev. 82.765

Number of obs: 960, groups: Item, 32; Subject, 30

**Fixed effects:**
- Interception: Estimate 2094.04, Std. Error 54.57, t value 38.37
- SentenceType1: Estimate 0.7154, Std. Error 0.35, t value 2.04
- SubjectTyp1: Estimate 1.7890, Std. Error 0.35, t value 5.06
- SentenceType1:SubjectTyp1: Estimate 188.75, Std. Error 17.71, t value 10.66

**Correlation of Fixed Effects:**
- (Intr) SntnT1 SbjcT1
- SentenceTyp1 0.023
- SubjectTyp1 -0.012 0.000
- SntncT1:ST1 0.000 -0.148 0.274

Intercept: Now mean RT overall

Significant overall RT difference for passive vs active sentences (across all subjects)

Significant overall RT difference for aphasics (across all sentence types)

Significant Special effect of aphasia + passive sentence

No correlation w/ intercept—we’ve successfully centered
Effects Coding: Sign Changes

- We picked one condition to be -0.5 and one to be 0.5
  - `contrasts(aphasia$SentenceType) = c(-0.5, 0.5)`
  - Here, Active was -0.5 and Passive was 0.5

- Should we worry that this affects our results?
  - Let’s try it the other way and see if we get something else
  - `contrasts(aphasia$SentenceType) = c(0.5, -0.5)`
  - Then, re-run the model
**Effects Coding: Sign Changes**

- Active is **-0.5**, Passive is **0.5**:  
  
<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>2094.04</td>
<td>54.57</td>
<td>38.37</td>
</tr>
<tr>
<td>SentenceType1</td>
<td>671.54</td>
<td>76.53</td>
<td>8.78</td>
</tr>
<tr>
<td>SubjectType1</td>
<td>178.90</td>
<td>78.32</td>
<td>2.28</td>
</tr>
<tr>
<td>SentenceType1:SubjectType1</td>
<td>188.75</td>
<td>17.71</td>
<td>10.66</td>
</tr>
</tbody>
</table>

  “RT 671 ms longer for Passive than for Active”

- Active is **0.5**, Passive is **-0.5**:  
  
<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>2094.04</td>
<td>54.57</td>
<td>38.37</td>
</tr>
<tr>
<td>SentenceType1</td>
<td>-671.54</td>
<td>76.53</td>
<td>-8.78</td>
</tr>
<tr>
<td>SubjectType1</td>
<td>178.90</td>
<td>78.32</td>
<td>2.28</td>
</tr>
<tr>
<td>SentenceType1:SubjectType1</td>
<td>-188.75</td>
<td>17.71</td>
<td>-10.66</td>
</tr>
</tbody>
</table>

  “RT 671 ms shorter for Active than for Passive”

- Flipping the signs of the code just changes the sign of the results  
  - Doesn’t affect absolute value or significance

- Choose whichever makes sense for your question:  
  - “Passive is slower than Active” vs “Active is faster than Passive”
**Effects Coding: Why -0.5 & 0.5?**

**PASSIVE**

1. Contrast Code
   - Passive: 0.5
   - Active: -0.5

1 unit change in contrast is the difference between sentence types

**ACTIVE**

1. Contrast Code
   - Passive: 1
   - Active: -1

1 unit change in contrast is only *half* the difference between levels
Effects Coding: Why -0.5 & 0.5?

- What if we used $(-1, 1)$ instead?
- Doesn't affect significance test
- Does make it harder to interpret the estimate
  - Parameter estimate is only half of the actual difference in means

![Table Comparison]

<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>2094.04</td>
<td>54.57</td>
<td>38.37</td>
</tr>
<tr>
<td>SubjectType1</td>
<td>178.90</td>
<td>78.32</td>
<td>2.28</td>
</tr>
<tr>
<td>SentenceType1</td>
<td>671.54</td>
<td>76.53</td>
<td>8.78</td>
</tr>
<tr>
<td>SubjectType1:SentenceType1</td>
<td>188.75</td>
<td>17.71</td>
<td>10.66</td>
</tr>
</tbody>
</table>

![Table Comparison]

<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>2094.039</td>
<td>54.572</td>
<td>38.37</td>
</tr>
<tr>
<td>SubjectType1</td>
<td>178.807</td>
<td>78.322</td>
<td>2.28</td>
</tr>
<tr>
<td>SentenceType1</td>
<td>335.771</td>
<td>38.263</td>
<td>8.78</td>
</tr>
<tr>
<td>SubjectType1:SentenceType1</td>
<td>94.376</td>
<td>8.854</td>
<td>10.66</td>
</tr>
</tbody>
</table>
Week 6: Coding Predictors I

- Centering Continuous Variables
- Introduction to Contrast Coding
- Dummy Coding
  - What it Means
  - How to Change Codes
  - Interactions
- Effects Coding
- Main Effects vs Simple Main Effects
- Unbalanced Factors
**Simple Main Effects vs Main Effects**

- **Treatment coding** and **effects coding** also change our interpretation of the non-intercept effects:

  - **Treatment coding** (of SentenceType):  
    - (Intercept)  
      - Estimate: 1716.01  
        - Std. Error: 76.85  
          - t value: 22.330
    - SubjectType1  
      - Estimate: 84.52  
        - Std. Error: 76.37  
          - t value: 1.107

  - Effect of SubjectType **within the baseline level** of SentenceType
    - “Simple main effect” or “simple effect” – not a “real” main effect

  - **Effects coding** (of SentenceType):  
    - (Intercept)  
      - Estimate: 2094.04  
        - Std. Error: 54.57  
          - t value: 38.37
    - SubjectType1  
      - Estimate: 178.90  
        - Std. Error: 78.32  
          - t value: 2.28

- Non-significant RT difference for aphasics (among active sentences)

- Significant RT difference for aphasics (across all sentence types)

- Overall effect of SubjectType averaged across sentence types
  - “Main effect”
Simple Main Effects vs Main Effects

• Again, both of these are, in principle, reasonable questions to ask…

• In factorial experiments, traditional to talk about the main effects averaged across other variables
  • “Main effect of aphasia,” “Overall effect of priming,” “Overall effect of study strategy,” “Main effect of ambiguity”…
  • If you want to talk about main effects in this way, don’t use treatment / dummy coding!

• In other designs, treatment coding may be the most appropriate!
Week 6: Coding Predictors I

- Centering Continuous Variables
- Introduction to Contrast Coding
- Dummy Coding
  - What it Means
  - How to Change Codes
  - Interactions
- Effects Coding
- Main Effects vs Simple Main Effects
- Unbalanced Factors
Unbalanced Factors

- Sometimes, we may have differing numbers of observations per level.

Possible reasons:

- Some categories naturally more common
  - e.g., college majors
- Categories may be equally common in the population, but we have sampling error
  - e.g., ended up 60% female participants, 40% male
- Study was designed so that some conditions are more common
  - e.g., more “control” subjects than “intervention” subjects
- We wanted equal numbers of observations, but there was an error
  - e.g., data loss due to computer problems
Weighted Coding

- “For the average student, does course size predict probability of graduation?”
  - Random sample of 200 Pitt undergrads
  - 5 are student athletes and 195 are not

- We could try to apply effects coding by centering around the mean and getting $(0.5, -0.5)$, but…
Weighted Coding

- An intercept at 0 would no longer correspond to the overall mean.
- As a scale, this would be totally unbalanced.
- To fix balance, we need to assign a heavier weight to Athlete.

But “not athlete” is actually far more common.
Weighted Coding

- Change codes so the mean is 0
  - \( c(.975, -0.25) \)

- centerfactor.R on CourseWeb will calculate this
Weighted Coding

• Change the codes so that the mean is 0 again
  • Used when the imbalance reflects something real
  • Like Type II sums of squares

• “For the average student, does course size predict graduation rates?”
  • Average student is not a student athlete, and our answer to the question about an “average student” should reflect this!
Unweighted Coding

• Oops! Our experiment loaded up the wrong image for one of our Passive sentences (“Groceries”)

• Now, imbalance is an accident and not meaningful
  • In fact, we’d like to get rid of it!

• Retain the (-0.5, 0.5) codes
  • Weights the two conditions equally—because the imbalance isn’t meaningful
  • Like Type III sums of squares
  • Probably what you want for factorial experiments
Unbalanced Factors: Summary

- **Weighted coding:** Change the codes so that the mean is 0
  - Use when the imbalance reflects something *real*
  - Can be done with `centerfactor.R` on CourseWeb

Mean across each individual:

- **Unweighted coding:** Keep the codes as -0.5 and 0.5
  - Use when the imbalance is an *accident* that we want to eliminate

Mean of the two levels:

(with perfectly balanced factors, these two methods are identical)