Week 7.2: Simple vs. Main Effects

- Follow-ups
  - Fixed Effect Correlation
    - ICC
- Categorical Variables with 2 Categories
  - Intro & Terminology
  - Treatment Coding
    - What it Does
    - How to Change Codes
    - Interactions
  - Effects Coding
    - One Variable
    - Two Variables
    - Sign & Scale Changes
- Simple Effects vs. Main Effects
- Unbalanced Factors
Follow-up #1: Fixed Effect Correlations

• What does this mean?

```plaintext
Fixed effects:
             Estimate Std. Error   df  t value Pr(>|t|)
(Intercept) 1690.421    152.547  24.800   11.081  4.28e-11 ***
NumDots      9.401       1.178   27.562   7.984  1.21e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
     (Intr)
NumDots -0.499
```
Follow-up #1: Fixed Effect Correlations

- Remember how we found our model results: Iterative estimation until convergence
- `verbose=2`

Each of these iterations represents a different set of estimates for our fixed effects:

```
Iteration: 1  
f(x) = 14044.221883
Iteration: 2  
f(x) = 14054.544644
Iteration: 3  
f(x) = 14261.542257
Iteration: 4  
f(x) = 14047.027132
Iteration: 5  
f(x) = 14044.635197
Iteration: 6  
f(x) = 14044.159759
Iteration: 7  
f(x) = 14044.119801
Iteration: 8  
f(x) = 14044.124273
Iteration: 9  
f(x) = 14044.119527
Iteration: 10 
f(x) = 14044.119526
Iteration: 11 
f(x) = 14044.119520
Iteration: 12 
f(x) = 14044.119520
Iteration: 13 
f(x) = 14044.119520
```
Follow-up #1: Fixed Effect Correlations

- Remember how we found our model results: Iterative estimation until convergence
- verbose=2

Each of these iterations represents a different set of estimates for our fixed effects

Sometimes each estimate may be bigger and sometimes smaller
Follow-up #1: Fixed Effect Correlations

• Remember how we found our model results: Iterative estimation until convergence

• verbose=2

Each of these iterations represents a different set of estimates for our fixed effects.

Sometimes each estimate may be bigger and sometimes smaller.

Fixed effect correlation is whether these two columns are correlated i.e., on iteration steps where we estimate a larger intercept, do we estimate a larger or smaller NumDots effect?
Follow-up #1: Fixed Effect Correlations

- What can we do with this information?
  - In most cases, very little
  - Very high correlations (e.g., ±.90) could be a sign of multicollinearity
    - Estimator can’t separate the two effects

- Can be used to verify centering worked
  - Correlations with intercept will be close to 0 IF you centered

<table>
<thead>
<tr>
<th></th>
<th>UNCENTERED</th>
<th>CENTERED</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed effects:</td>
<td>Estimate Std. Error    df  t value Pr(&gt;</td>
<td>t</td>
</tr>
<tr>
<td>(Intercept)</td>
<td>1600.421 152.547 24.800 11.081 4.28e-11 ***</td>
<td>(Intercept) 2323.448 127.129 28.991 18.276 &lt; 2e-16 ***</td>
</tr>
<tr>
<td></td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Signif. codes:</td>
<td>0 ‘<strong><strong>’ 0.001 ‘</strong>*’ 0.01 ‘</strong>’ 0.05 ‘.’ 0.1 ‘ ’ 1</td>
<td>0 ‘<strong><strong>’ 0.001 ‘</strong>*’ 0.01 ‘</strong>’ 0.05 ‘.’ 0.1 ‘ ’ 1</td>
</tr>
<tr>
<td>Correlation of Fixed Effects: (Intr)</td>
<td>NumDots -0.499</td>
<td>Correlation of Fixed Effects: (Intr)</td>
</tr>
</tbody>
</table>
Week 7.2: Simple vs. Main Effects

- Follow-ups
- Fixed Effect Correlation
- ICC
- Categorical Variables with 2 Categories
  - Intro & Terminology
  - Treatment Coding
    - What it Does
    - How to Change Codes
    - Interactions
  - Effects Coding
    - One Variable
    - Two Variables
    - Sign & Scale Changes
- Simple Effects vs. Main Effects
- Unbalanced Factors
Follow-up #2: ICC

- One way we can assess the importance of our clustering variables is with the *intra-class correlation coefficient* (ICC).

- Remember our math data?

LEVEL 3
Sampled SCHOOLS

LEVEL 2
Sampled CLASSROOMS

LEVEL 1
Sampled STUDENTS
Follow-up #2: ICC

• 2 clustering variables here:
  • Classrooms
  • Schools

• How much of the total variance do these clustering variables account for?
  • We want to look the raw variance, not “explained” by any fixed effect

• Empty Means model:
  • `empty.means.model <- lmer(FinalMathScore ~ 1 + (1|Classroom) + (1|School), data=math)`
**Follow-up #2: ICC**

- **Empty Means results:**

  ```
  Random effects:
  Groups       Name  Variance Std.Dev.
  Classroom    (Intercept)  6.240  2.498
  School       (Intercept)  1.126  1.061
  Residual      43.536  6.598
  Number of obs: 720, groups: Classroom, 24; School, 3
  ```

  3 sources of variance present in the data. Which is most important?

  \[
  ICC = \frac{\text{Variance of Random Effect of Interest}}{\text{Sum of All Random Effect Variances}} = \frac{\text{Classroom Variance}}{\text{Classroom Variance} + \text{School Variance} + \text{Residual Variance}}
  \]

- **library(performance)**
- **icc(empty.means.model, by_group=TRUE)**

  12% of the variance can be attributed to Classroom differences
  2% of the variance can be attributed to School differences
  The remaining 86% is at the student level (residual)
Follow-up #2: ICC

- Empty Means results:

<table>
<thead>
<tr>
<th>Groups</th>
<th>Name</th>
<th>Variance</th>
<th>Std.Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classroom</td>
<td>(Intercept)</td>
<td>6.240</td>
<td>2.498</td>
</tr>
<tr>
<td>School</td>
<td>(Intercept)</td>
<td>1.126</td>
<td>1.061</td>
</tr>
<tr>
<td>Residual</td>
<td></td>
<td>43.536</td>
<td>6.598</td>
</tr>
</tbody>
</table>

Number of obs: 720, groups: Classroom, 24; School, 3

3 sources of variance present in the data. Which is most important?

\[
\text{ICC} = \frac{\text{Variance of Random Effect of Interest}}{\text{Sum of All Random Effect Variances}} = \frac{\text{Classroom Variance}}{\text{Classroom Variance} + \text{School Variance} + \text{Residual Variance}}
\]

- `library(performance)` (must install the package first)
- `icc(empty.means.model, by_group=FALSE)`

Total explained by all random effects added together (Classroom + School) = 14.5%
Follow-up #2: ICC

- Empty Means results:

<table>
<thead>
<tr>
<th>Random effects:</th>
<th>Variance</th>
<th>Std.Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Groups Name</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Classroom (Intercept)</td>
<td>6.240</td>
<td>2.498</td>
</tr>
<tr>
<td>School (Intercept)</td>
<td>1.126</td>
<td>1.061</td>
</tr>
<tr>
<td>Residual</td>
<td>43.536</td>
<td>6.598</td>
</tr>
<tr>
<td>Number of obs:</td>
<td>720,</td>
<td>Classroom, 24; School, 3</td>
</tr>
</tbody>
</table>

ICC = \[
\frac{\text{Variance of Random Effect of Interest}}{\text{Sum of All Random Effect Variances}} = \frac{\text{Classroom Variance}}{\text{Classroom Variance} + \text{School Variance} + \text{Residual Variance}}
\]

- Likelihood-ratio tests tell us whether random effects significantly contribute to the model (yes/no)
- ICC tells us about the relative importance (≈ “effect size”)
Week 7.2: Simple vs. Main Effects

- Follow-ups
  - Fixed Effect Correlation
  - ICC
- Categorical Variables with 2 Categories
  - Intro & Terminology
    - Treatment Coding
      - What it Does
      - How to Change Codes
      - Interactions
    - Effects Coding
      - One Variable
      - Two Variables
      - Sign & Scale Changes
- Simple Effects vs. Main Effects
- Unbalanced Factors
Today’s Dataset

- Task: Decide whether a picture matches a sentence; measure RT
Today’s Dataset

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

“The dog was chased by the man.”
Today’s Dataset

- Task: Decide whether a picture matches a sentence; measure RT
- Each Item: Unique sentence w/ a unique picture
  - No picture or sentence repeats

“The dog was chased by the man.”

“The bee stung the man.”
Today’s Dataset

- Task: Decide whether a picture matches a sentence; measure RT
- Each Item: Unique sentence w/ a unique picture
- 16 people with aphasia and 16 healthy controls (SubjectType)
- All participants see the same sentences, which vary in SentenceType (Active or Passive)
  - Active (more common): “Dog bites man.”
    - Subject of the sentence is performing the action
  - Passive: “The dog was bitten by the man.”
    - Subject of the sentence is having an action performed to them
Today’s Dataset

- Task: Decide whether a picture matches a sentence; measure RT
- Each Item: Unique sentence w/ a unique picture
- 16 people with aphasia and 16 healthy controls (SubjectType)
- All participants see the same sentences, which vary in SentenceType (Active or Passive)
  - Active (more common): “Dog bites man.”
  - Passive: “The dog was bitten by the man.”
- Which fixed effect(s) are between-subjects?
- Which fixed effect(s) are within-subjects?
  - Hint: Imagine we had only 1 subject. If we could still test the effect of a variable, it’s within-subjects.
Today’s Dataset

- Task: Decide whether a picture matches a sentence; measure RT
- Each **Item**: Unique sentence w/ a unique picture
- 16 people with aphasia and 16 healthy controls (**SubjectType**)
- All participants see the same sentences, which vary in **SentenceType** (Active or Passive)
  - Active (more common): “Dog bites man.”
  - Passive: “The dog was bitten by the man.”
- Which fixed effect(s) are **between**-subjects?
  - **SubjectType**
- Which fixed effect(s) are **within**-subjects?
  - **SentenceType**
Today’s Dataset

- Task: Decide whether a picture matches a sentence; measure RT
- Each Item: Unique sentence w/ a unique picture
- 16 people with aphasia and 16 healthy controls (SubjectType)
- All participants see the same sentences, which vary in SentenceType (Active or Passive)
  - Active (more common): “Dog bites man.”
  - Passive: “The dog was bitten by the man.”

- Which fixed effect(s) are between-items?
- Which fixed effect(s) are within-items?
  - Hint: Imagine we had only 1 sentence. If we could still test the effect of a variable, it’s within-items.
Today’s Dataset

- Task: Decide whether a picture matches a sentence; measure RT
- Each Item: Unique sentence w/ a unique picture
- 16 people with aphasia and 16 healthy controls (SubjectType)
- All participants see the same sentences, which vary in SentenceType (Active or Passive)
  - Active (more common): “Dog bites man.”
  - Passive: “The dog was bitten by the man.”
- Which fixed effect(s) are between-items?
  - SentenceType
- Which fixed effect(s) are within-items?
  - SubjectType
The Big Picture

- Both SentenceType and SubjectType are categorical variables

- When categorical variables are independent / predictor variables, need to make choices about how to code them
  - 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
Terminology

• **Factor**: A variable with a fixed set of categories
  • `as.factor()` tells R to treat a variable as categorical
  • Displays frequency counts in `summary()`

```r
aphasia %>%
  mutate(SentenceType = as.factor(SentenceType),
         SubjectType = as.factor(SubjectType)) -> aphasia
```

SentenceType
- Active: 480
- Passive: 480
Terminology

- **Factor**: A variable with a fixed set of categories
  - `as.factor()` tells R to treat a variable as categorical
  - Displays frequency counts in `summary()`

- **Levels**: The individual categories within a factor
  - “Active” versus “Passive”
  - “Aphasia” versus “Healthy control”
  - …whether experimental or observational
  - Today, we will look at cases where the factor has 2 levels
Terminology

- **Factorial Design**: A design where each combination of levels appears
- Common in experimental (& quasi-experimental) contexts where we manipulate each variable

<table>
<thead>
<tr>
<th>SubjectType</th>
<th>SentenceType</th>
<th>ACTIVE, APHASIA</th>
<th>240 observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aphasia</td>
<td>Active</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aphasia</td>
<td>Passive</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>Active</td>
<td>ACTIVE, CONTROL</td>
<td>240 observations</td>
</tr>
<tr>
<td>Control</td>
<td>Passive</td>
<td>PASSIVE, CONTROL</td>
<td>240 observations</td>
</tr>
</tbody>
</table>
**Terminology**

- **Factorial Design**: A design where each combination of levels appears.
- Common in experimental (& quasi-experimental) contexts where we manipulate each variable.

- **Cell**: One individual combination.

<table>
<thead>
<tr>
<th>SubjectType</th>
<th>SentenceType</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aphasia</td>
<td>Active, Aphasia</td>
</tr>
<tr>
<td>Aphasia</td>
<td>Passive, Aphasia</td>
</tr>
<tr>
<td>Control</td>
<td>Active, Control</td>
</tr>
<tr>
<td>Control</td>
<td>Passive, Control</td>
</tr>
</tbody>
</table>
Week 7.2: Simple vs. Main Effects

- Follow-ups
  - Fixed Effect Correlation
  - ICC
- Categorical Variables with 2 Categories
  - Intro & Terminology
    - Treatment Coding
      - What it Does
        - How to Change Codes
        - Interactions
    - Effects Coding
      - One Variable
      - Two Variables
      - Sign & Scale Changes
- Simple Effects vs. Main Effects
- Unbalanced Factors
Introduction to Contrast Coding

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

\[ \text{RT} = \text{Intercept} + \gamma_{100} \times \text{NumDots} + \gamma_{200} \times \text{MathAnxiety} \]

- But what about active vs passive sentence?

\[ \text{RT} = \text{Intercept} + \gamma_{100} \times \text{NP} \]
Introduction to Contrast Coding

• But what about active vs passive sentence?

• 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

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

Active sentence: 0
Passive sentence: 1
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
  • Remember, today we are just looking at factors with 2 levels
• Default ordering is alphabetical: First level is 0, second is 1
  • We’ll see how to change this soon
• contrasts(aphasia$SentenceType)

<table>
<thead>
<tr>
<th></th>
<th>Passive</th>
<th>Active</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Passive</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

Active coded as 0
Passive coded as 1
Let's do a model that just examines the effect of sentence type in this task:

```r
SentenceTypeModel <- lmer(RT ~ 1 + SentenceType +
SUBJECT RANDOM EFFECTS
ITEM RANDOM EFFECTS
data = aphasia)
```

*Hint*: `SentenceType` varies within-subjects, but only between items
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:

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

  Passive Sentences
  \[ RT = \gamma_{000} + \gamma_{100} \times \text{SentenceType} \]
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**
  \[ RT = \gamma_{000} + \gamma_{100} \]
  \[ \begin{array}{c} 0 \\ \end{array} \]

- **Passive Sentences**
  \[ RT = \gamma_{000} + \gamma_{100} \]
  \[ \begin{array}{c} 1 \\ \end{array} \]
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 \( (\gamma_{000}) \) is just the mean RT for active sentences

- **Passive Sentences**
  \[ RT = \gamma_{000} + \gamma_{100} \]
  
  Intercept \( (\gamma_{000}) \) is just the mean RT for active sentences
Treatment Coding (Dummy Coding)

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

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

Intercept \((\gamma_{000})\) is just the mean RT for active sentences.

What is the difference between the equations for the two sentence types?

SentenceType effect \((\gamma_{100})\) is the difference in RT between passive & active sentences.

Active Sentences

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

Passive Sentences

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

Intercept ($\gamma_{000}$): RT for active sentences is 1758 ms

Sentence Type ($\gamma_{100}$): Passive sentences take an additional 672 ms to read

- Treatment coding makes one level the baseline and compares everything to that
Week 7.2: Simple vs. Main Effects

- Follow-ups
- Fixed Effect Correlation
- ICC
- Categorical Variables with 2 Categories
- Intro & Terminology
  - Treatment Coding
    - What it Does
    - How to Change Codes
      - Interactions
  - Effects Coding
    - One Variable
    - Two Variables
    - Sign & Scale Changes
- Simple Effects vs. Main Effects
- Unbalanced Factors
Changing Codes

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

  ![Contrast Table]
  
  Control  | 0
  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 7.2: Simple vs. Main Effects

- Follow-ups
  - Fixed Effect Correlation
  - ICC
- Categorical Variables with 2 Categories
  - Intro & Terminology
    - Treatment Coding
      - What it Does
      - How to Change Codes
      - Interactions
    - Effects Coding
      - One Variable
      - Two Variables
      - Sign & Scale Changes
- Simple Effects vs. Main Effects
- Unbalanced Factors
Treatment Coding: Two Variables

- Now, we’d like SentenceType and SubjectType to interact:
- `Model.Maximal <- lmer(RT ~ 1 + SentenceType * SubjectType + SUBJECT RANDOM EFFECTS + ITEM RANDOM EFFECTS, data = aphasia)`

- **Hint #1:** Remember that we can include a random slope by subjects for within-subjects variables but not for between-subjects variables.

- **Hint #2:** Does each subject see more than one SentenceType? Is each subject more than one SubjectType?
Treatment Coding: Two Variables

• Now, we’d like SentenceType and SubjectType to interact:
• Model.Maximal <- lmer(RT ~
  1 + SentenceType * SubjectType +
  (1 + SentenceType | Subject) +
  (1 + SubjectType | Item),
data = aphasia)

• Hint #1: Remember that we can include a random slope by items for within-items variables but not for between-items variables

• Hint #2: Is each item presented as more than one SentenceType? Is each item presented to more than one SubjectType?
**Treatment Coding: Two Variables**

- Now, we’d like SentenceType and SubjectType to interact:
- `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:

  - Active, Control Subj.
  - Passive, Control Subj.
  - Active, Aphasics
  - Passive, Aphasics

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

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

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

RT = $\gamma_{000} + \gamma_{100}\text{SentenceType} + \gamma_{200}\text{SubjectType} + \gamma_{1200}\text{SentenceTypeSubjectType}$
Treatment Coding: Two Variables

- Our design now has four cells:

  \[
  \text{Active, Control Subj.:} \quad RT = \gamma_{000} + \gamma_{100} \text{SentenceType} + \gamma_{200} \text{SubjectType} + \gamma_{1200} \text{SentenceTypeSubjectType}
  \]

  \[
  \text{Passive, Control Subj.:} \quad RT = \gamma_{000} + \gamma_{100} \text{SentenceType} + \gamma_{200} \text{SubjectType} + \gamma_{1200} \text{SentenceTypeSubjectType}
  \]

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

  \[
  \text{Passive, Aphasics} \quad RT = \gamma_{000} + \gamma_{100} \text{SentenceType} + \gamma_{200} \text{SubjectType} + \gamma_{1200} \text{SentenceTypeSubjectType}
  \]
Treatment Coding: Two Variables

• Our design now has four cells:

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

Passive, Control Subj.  
RT = \gamma_{000} + \gamma_{100} + \gamma_{200} + \gamma_{1200} = 1 + 0 + 0 = 1

Active, Aphasic  
RT = \gamma_{000} + \gamma_{100} + \gamma_{200} + \gamma_{1200} = 1 + 0 + 0 = 1

Passive, Aphasic  
RT = \gamma_{000} + \gamma_{100} + \gamma_{200} + \gamma_{1200} = 1 + 0 + 0 = 1
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**
    \[ RT = \gamma_{000} + \gamma_{100} + \gamma_{200} \]

  - **Passive, Aphasics**
    \[ RT = \gamma_{000} + \gamma_{100} + \gamma_{200} + \gamma_{1200} \]

Intercept (\( \gamma_{000} \)) is the RT when all variables at their baseline: active sentence type, healthy control subject

SentenceType (\( \gamma_{100} \)) : Passive vs active difference for baseline healthy controls

\( \gamma_{200} \)

SentenceTypeSubjectType

SubjectType

SentenceTypeSubjectType

0 1
Treatment Coding: Two Variables

- Our design now has four cells:

Active, Control Subj.

- $\text{RT} = \gamma_{000}$
- Intercept ($\gamma_{000}$) is the RT when all variables at their baseline: active sentence type, healthy control subject

Passive, Control Subj.

- $\text{RT} = \gamma_{000} + \gamma_{100}$
- SentenceType ($\gamma_{100}$): Passive vs active difference for baseline healthy controls

Active, Aphasics

- $\text{RT} = \gamma_{000} + \gamma_{200}$
- SubjectType ($\gamma_{200}$): Aphasia vs control difference for baseline active sentences

Passive, Aphasics

- $\text{RT} = \gamma_{000} + \gamma_{100} \text{SentenceType} + \gamma_{200} \text{SubjectType} + \gamma_{1200} \text{SentenceTypeSubjectType}$
Treatment Coding: Two Variables

- Our design now has four cells:

<table>
<thead>
<tr>
<th>Conditions</th>
<th>Equation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active, Control Subj.</td>
<td>$RT = \gamma_{000}$</td>
<td>Intercept ($\gamma_{000}$) 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>SentenceType ($\gamma_{100}$): Passive vs active difference for baseline healthy controls</td>
</tr>
<tr>
<td>Active, Aphasics</td>
<td>$RT = \gamma_{000} + \gamma_{200}$</td>
<td>SubjectType ($\gamma_{200}$): Aphasia vs control difference for baseline active sentences</td>
</tr>
<tr>
<td>Passive, Aphasics</td>
<td>$RT = \gamma_{000} + \gamma_{100} + \gamma_{200} + 1$</td>
<td>If no special effect of passive sentence and aphasia, we’d just have these two effects</td>
</tr>
</tbody>
</table>
Treatment Coding: Two Variables

Our design now has four cells:

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

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

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

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

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>Condition</th>
<th>Equation</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active, Control Subj.</td>
<td>( \text{RT} = \gamma_{000} )</td>
<td>Intercept (( \gamma_{000} )) is the RT when all variables at their baseline: active sentence type, healthy control subject</td>
</tr>
<tr>
<td>Passive, Control Subj.</td>
<td>( \text{RT} = \gamma_{000} + \gamma_{100} )</td>
<td>SentenceType (( \gamma_{100} )) : Passive vs active difference for baseline healthy controls</td>
</tr>
<tr>
<td>Active, Aphasics</td>
<td>( \text{RT} = \gamma_{000} + \gamma_{200} )</td>
<td>SubjectType (( \gamma_{200} )) : Aphasia vs control difference for baseline active sentences</td>
</tr>
<tr>
<td>Passive, Aphasics</td>
<td>( \text{RT} = \gamma_{000} + \gamma_{100} + \gamma_{200} + \gamma_{1200} )</td>
<td></td>
</tr>
</tbody>
</table>
## Treatment Coding: Two Variables

- Our design now has four cells:

<table>
<thead>
<tr>
<th>Condition</th>
<th>RT Equation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active, Control Subj.</td>
<td>$RT = \gamma_{000}$</td>
<td>Intercept ($\gamma_{000}$) 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>SentenceType ($\gamma_{100}$): Passive vs active difference for baseline healthy controls</td>
</tr>
<tr>
<td>Active, Aphasics</td>
<td>$RT = \gamma_{000} + \gamma_{200}$</td>
<td>SubjectType ($\gamma_{200}$): Aphasia vs control difference for baseline active sentences</td>
</tr>
<tr>
<td>Passive, Aphasics</td>
<td>$RT = \gamma_{000} + \gamma_{100} + \gamma_{200} + \gamma_{1200}$</td>
<td>Interaction: Special effect of aphasia and passive sentence</td>
</tr>
</tbody>
</table>
Treatment Coding: Model Results

Intercept: RT for healthy controls, active 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:**
- **Groups**
  - Item: (Intercept) 47858.72 218.766
  - SubjectTypeAphasia: 55.88 7.476 -1.00
- **Subject**
  - (Intercept) 43288.90 208.060
  - SentenceTypePassive: 1443.32 37.991 0.27
- Residual: 6850.05 82.765

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

**Fixed effects:**

|             | Estimate | Std. Error | df  | t value | Pr(>|t|) |
|-------------|----------|------------|-----|---------|----------|
| (Intercept) | 1716.01  | 76.85      | 57.43 | 22.330  | < 2e-16  | ***     |
| SentenceTypePassive | 577.17 | 78.33 | 30.94 | 7.368  | 2.73e-08 | ***     |
| SubjectTypeAphasia | 84.52 | 76.37 | 28.03 | 1.107  | 0.278    |
| SentenceTypePassive:SubjectTypeAphasia | 188.75 | 17.71 | 28.82 | 10.659  | 1.64e-11 | ***     |

---

Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
- (Inter) SntncTP SbjcTA
- SntncTypPss -0.478
- SbectTypAph -0.514 -0.002
- SntncTP:STA -0.008 -0.258 0.165

**Intercept:** RT for healthy controls, active 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**
Week 7.2: Simple vs. Main Effects

- Follow-ups
  - Fixed Effect Correlation
  - ICC

- Categorical Variables with 2 Categories
  - Intro & Terminology
  - Treatment Coding
    - What it Does
    - How to Change Codes
    - Interactions
  - Effects Coding
    - One Variable
      - Two Variables
      - Sign & Scale Changes
  - Simple Effects vs. 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

<table>
<thead>
<tr>
<th>Active</th>
<th>0</th>
<th>Subtract 0.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Passive</td>
<td>1</td>
<td>Subtract 0.5</td>
</tr>
</tbody>
</table>

\[ \text{Active} : 0.5, \quad \text{Passive} : 1 \]
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: \[
RT = \gamma_{000} + \gamma_{100}
\]

Imagine 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:

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

Active Sentences

\[
\text{RT} = \gamma_{000} + \gamma_{100} -0.5
\]

Passive Sentences

\[
\text{RT} = \gamma_{000} + \gamma_{100} 0.5
\]

The equations differ by 1 \(\gamma_{100}\). SentenceType effect (\(\gamma_{100}\)) is (still) the difference between conditions.
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

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

Intercept is what you expect when the other variables are 0. Where does a value of \( x=0 \) fall in this coding scheme?
Effects Coding (Sum Coding)

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

\[
RT = \gamma_000 + \gamma_{100} \times \text{SentenceType}
\]

Intercept is what you expect when the other variables are 0. Where does a value of \( x=0 \) fall in this coding scheme?

0 is midway between Active and Passive; i.e., the average of the conditions.
**Effects Coding (Sum Coding)**

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

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

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

    Sentence Type effect ($\gamma_{100}$) is (still) the difference between conditions

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

    Intercept ($\gamma_{000}$) is now the mean RT across all conditions.
Effects Coding (Sum Coding)

- Let’s first go back to our model with only SentenceType and try it with effects coding.

Old codes:

<table>
<thead>
<tr>
<th>SentenceType</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active</td>
<td>0</td>
</tr>
<tr>
<td>Passive</td>
<td>1</td>
</tr>
</tbody>
</table>

New codes:

- `contrasts(aphasia$SentenceType) <- c(-0.5, 0.5)`

<table>
<thead>
<tr>
<th>SentenceType</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active</td>
<td>-0.5</td>
</tr>
<tr>
<td>Passive</td>
<td>0.5</td>
</tr>
</tbody>
</table>
### Effects Coding Results

#### TREATMENT CODING

**Different Intercepts (Active baseline vs. mean)**

**Same SentenceType effect**

| Fixed effects | Estimate | Std. Error | df | t value | Pr(>|t|) |
|---------------|----------|------------|----|---------|----------|
| (Intercept)   | 1758.27  | 66.02      | 53.42 | 26.631  | < 2e-16 *** |
| SentenceTypePassive | 671.54  | 78.50 | 33.60 | 8.555 | 5.97e-10 *** |

---

Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:

(Intr)

SntncTypPss -0.521

---

| Fixed effects | Estimate | Std. Error | df | t value | Pr(>|t|) |
|---------------|----------|------------|----|---------|----------|
| (Intercept)   | 2094.04  | 56.58      | 58.01 | 37.011  | < 2e-16 *** |
| SentenceType1 | 671.54  | 78.49 | 33.60 | 8.556 | 5.95e-10 *** |

---

Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:

(Intr)

SentencTyp1 0.086
Week 7.2: Simple vs. Main Effects

- Follow-ups
  - Fixed Effect Correlation
  - ICC
- Categorical Variables with 2 Categories
  - Intro & Terminology
  - Treatment Coding
    - What it Does
    - How to Change Codes
    - Interactions
- Effects Coding
  - One Variable
  - Two Variables
    - Sign & Scale Changes
- Simple Effects vs. Main Effects
- Unbalanced Factors
Effects Coding (Sum Coding)

• Now let’s look at both variables together:
• Old codes:

<table>
<thead>
<tr>
<th>SENTENCETYPE</th>
<th>SUBJECTTYPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active</td>
<td>Aphasia</td>
</tr>
<tr>
<td>Passive</td>
<td>Control</td>
</tr>
</tbody>
</table>

| 0             | 1            |
| 1             | 0            |

• New codes:
  • contrasts(aphasia$SentenceType) <- c(-0.5,0.5)
  • contrasts(aphasia$SubjectType) <- c(0.5,-0.5)
Effects Coding: Full Model Results

- Intercept: Now mean RT overall
- Significant overall RT difference for passive vs active sentences (across all subject types)
- 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
Week 7.2: Simple vs. Main Effects

- Follow-ups
  - Fixed Effect Correlation
  - ICC
- Categorical Variables with 2 Categories
  - Intro & Terminology
  - Treatment Coding
    - What it Does
    - How to Change Codes
    - Interactions
  - Effects Coding
    - One Variable
    - Two Variables
    - Sign & Scale Changes
- Simple Effects vs. Main Effects
- Unbalanced Factors
Effects Coding: Sign Changes

Hang on! You coded SentenceType by making active sentences -0.5 and passive sentences 0.5. How do I know that’s the right way to do it?

• 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:**
  
  |                | Estimate | Std. Error | df | t value | Pr(>|t|) |
  |----------------|----------|------------|----|---------|----------|
  | (Intercept)    | 2094.04  | 54.57      | 57.50 | 38.372  | < 2e-16  ***|
  | SentenceType1  | 671.54   | 76.53      | 30.49 | 8.775   | 7.60e-10 ***|
  | SubjectType1   | 178.90   | 78.32      | 28.02 | 2.284   | 0.0301   * |
  | SentenceType1:SubjectType1 | 188.75   | 17.71      | 28.82 | 10.659  | 1.64e-11 ***|

  “RT 671 ms longer for Passive than for Active”

- **Active is 0.5, Passive is -0.5:**
  
  |                | Estimate | Std. Error | df | t value | Pr(>|t|) |
  |----------------|----------|------------|----|---------|----------|
  | (Intercept)    | 2094.04  | 54.57      | 57.50 | 38.372  | < 2e-16  ***|
  | SentenceType1  | -671.54  | 76.53      | 30.49 | -8.775  | 7.60e-10 ***|
  | SubjectType1   | 178.90   | 78.32      | 28.02 | 2.284   | 0.0301   * |
  | SentenceType1:SubjectType1 | -188.75  | 17.71      | 28.82 | -10.659 | 1.64e-11 ***|

  “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

CONTRAST CODE

PASSIVE

CONTRAST CODE

ACTIVE

1 unit change in contrast IS the difference between sentence types

ACTIVE

1 unit change in contrast IS only half the difference between levels

1

-0.5

1

-1

1

2

}
**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

---

**SENTENCE TYPE:**
\(c(-0.5, 0.5)\)

**Fixed effects:**

|              | Estimate | Std. Error | df  | t value | Pr(>|t|) |
|--------------|----------|------------|-----|---------|----------|
| (Intercept)  | 2094.04  | 54.57      | 57.50| 38.372  | < 2e-16  *** |
| SentenceType1| 671.54   | 76.53      | 30.49| 8.775   | 7.60e-10 *** |
| SubjectType1 | 178.90   | 78.32      | 28.02| 2.284   | 0.0301 *  |
| SentenceType1:SubjectType1 | 188.75 | 17.71      | 28.82| 10.659  | 1.64e-11 *** |

**SENTENCE TYPE:**
\(c(-1, 1)\)

**Fixed effects:**

|              | Estimate | Std. Error | df  | t value | Pr(>|t|) |
|--------------|----------|------------|-----|---------|----------|
| (Intercept)  | 2094.039 | 54.571     | 57.50| 38.373  | < 2e-16  *** |
| SentenceType1| 335.771  | 38.263     | 30.49| 8.775   | 7.59e-10 *** |
| SubjectType1 | 178.897  | 78.321     | 28.017| 2.284   | 0.0301 *  |
| SentenceType1:SubjectType1 | 94.376 | 8.855      | 28.818| 10.658  | 1.64e-11 *** |
Week 7.2: Simple vs. Main Effects

- Follow-ups
  - Fixed Effect Correlation
  - ICC
- Categorical Variables with 2 Categories
  - Intro & Terminology
  - Treatment Coding
    - What it Does
    - How to Change Codes
    - Interactions
  - Effects Coding
    - One Variable
    - Two Variables
    - Sign & Scale Changes
- Simple Effects vs. Main Effects
- Unbalanced Factors
Simple vs. Main Effects

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

- **Treatment coding** (of SentenceType):
  - Effect of SubjectType within the baseline level of SentenceType
    - “Simple effect” – not a “real” main effect

- **Effects coding** (of SentenceType):
  - Overall effect of SubjectType averaged across sentence types
    - “Main effect”

| Estimate | Std. Error | df | t value | Pr(>|t|) |
|----------|------------|----|---------|----------|
| (Intercept) | 1716.01 | 76.85 | 57.43 | 22.330 | < 2e-16 *** |
| SubjectType1 | 84.52 | 76.37 | 28.03 | 1.107 | 0.278 |

| Estimate | Std. Error | df | t value | Pr(>|t|) |
|----------|------------|----|---------|----------|
| (Intercept) | 2004.59 | 67.53 | 50.56 | 29.683 | < 2e-16 *** |
| SubjectType1 | 178.90 | 78.31 | 28.03 | 2.284 | 0.0301 * |

Non-significant RT difference for people with aphasia (among active sentences)

Significant RT difference for people with aphasia (across all sentence types)
Simple vs. Main Effects

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

• In factorial designs, 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, use effects coding, not treatment / dummy coding!

• In other designs, treatment coding may be the most appropriate!
Week 7.2: Simple vs. Main Effects

- Follow-ups
  - Fixed Effect Correlation
  - ICC
- Categorical Variables with 2 Categories
  - Intro & Terminology
  - Treatment Coding
    - What it Does
    - How to Change Codes
    - Interactions
- Effects Coding
  - One Variable
  - Two Variables
  - Sign & Scale Changes
- Simple Effects vs. 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 lost some because of **errors** or **exclusion criteria**
    - e.g., data loss due to computer problems
    - Dropping subjects below a minimum level of performance
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

• How can we make the intercept reflect the “average student”?
  • We could try to apply effects coding to the StudentAthlete variable 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, -.025)`
• `contr.helmert.weighted()` function in my `psycholing` package will calculate this
Weighted Coding

- **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")
  • It may have been sabotaged

• `UsableItem` column is **No** for this item

• First, can we remove this from our data?
• Some possibilities:
  • `aphasia %>% filter(UsableItem == 'Yes') -> aphasia`
  • `aphasia %>% filter(UsableItem != 'No') -> aphasia2`
  • etc.
Unweighted Coding

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

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

```
> summary(aphasia2)
  Subject  Item       SubjectType  SentenceType
S1       : 31  Astronaut: 30  Aphasia:465  Active:480
S11      : 31  Boy        : 30
S12      : 31  Breakfast: 30
S13      : 31  Burglar    : 30
S14      : 31  Cheese     : 30
(Other):744  (Other):750
```
Unweighted Coding

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

• Now, there’s an imbalance, but it’s 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 `contr.helmert.weighted()`

![Mean across each individual:](image)

- 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:](image)

- With balanced factors, these are identical