Course Business

• Clarification of course schedule:
  • Nov. 11, Nov. 18: Remaining lectures
  • Nov. 25, Dec. 2: Project presentations
• No papers due until Dec 2

• Final project presentation info. will be posted on CourseWeb at 5:00
  • Submit assignment on CourseWeb just by stating your preferred presentation date
  • Assignment will be graded & returned after the presentation

• This class is not a substitute for Psych regression class
Effect Sizes!

• Simplest measure: Parameter estimates
  • Effect of 1-unit change in predictor on outcome variable
  • “On average, increasing word length by 1 letter increased reading time by 7.62 ms” (see below)

```
Fixed effects:          Estimate  Std. Error  t value
                        (Intercept)  340.283      5.443     62.5
LENGTH                  7.615       0.675      11.3
LINESTARTYs 5            49.331      3.056      16.1
```

• “Personalized math problems increased odds of passing exam by 1.3 times.” [data not shown]
• Concrete!

• Sometimes variables are measured in different units
  • If you need to compare them directly, could standardize (z-score) first
  • “Effect on high school graduation rate of a 1-standard deviation difference in family income vs. a 1-standard deviation in class size.”
**Variance Explained**

- \( R^2: \text{cor} (\text{fitted}(\text{model1}), \text{dataframe}$\text{DependentMeasure})^2 \)
- But, this includes what’s predicted on basis of random effects like subjects & items
  - Saying “some subjects are faster than others” or “some schools are better than others” isn’t really “explaining” in a theoretically interesting way
- Compare to the \( R^2 \) of a model with *just* the subjects & items and no fixed effects
Effect Sizes!

- What’s a “large” effect?
  - Interpret in context of other effects in the domain, and…
  - Interventions: …cost, difficulty of implementation, etc.
    - E.g. aspirin: SMALL effect on heart disease, but used anyway because it’s cheap & easy to implement
  - Basic science: …predictions of competing theories
    - Question often isn’t “Is this a large effect?” but whether the data pattern matches the predictions of theory/theories being tested
Another Look at Nested Models

Hierarchical linear model

Level-1 Model: $Y_{ijk} = \pi_{0jk} + \pi_{1jk}(\text{Prior Achievement}_{ijk} - \bar{X}_1) + \pi_{2jk}(\text{Gender}_{ijk} - \bar{X}_2) + \pi_{3jk}(\text{Black}_{ijk} - \bar{X}_3) + \pi_{4jk}(\text{Hispanic}_{ijk} - \bar{X}_4) + \pi_{5jk}(\text{Native American}_{ijk} - \bar{X}_5) + \pi_{6jk}(\text{Other Ethnicity}_{ijk} - \bar{X}_6) + r_{ijk}$

Level-2 Model: $\pi_{0jk} = \beta_{00k} + \beta_{01k}\text{NSF}_{jk} + r_{0jk}$

Level-3 Model: $\beta_{00k} = \gamma_{000} + u_{00k}$

DV: Math score of individual student $i$

DV: Effect of class $j$

DV: Effect of school $k$
Another Look at Nested Models

Effect of being in classroom $j$ is a function of school effect ($\beta_{00k}$) + NSF variable + random class intercept $r_{0jk}$

Level-1 Model: $Y_{ijk} = \pi_{0jk} + \pi_{1jk}(\text{Prior Achievement}_{ijk} - \bar{X}_1) + \pi_{2jk}(\text{Gender}_{ijk} - \bar{X}_2) + \pi_{3jk}(\text{Black}_{ijk} - \bar{X}_3) + \pi_{4jk}(\text{Hispanic}_{ijk} - \bar{X}_4) + \pi_{5jk}(\text{Native American}_{ijk} - \bar{X}_5) + \pi_{6jk}(\text{Other Ethnicity}_{ijk} - \bar{X}_6) + r_{ijk}$

Level-2 Model: $\pi_{0jk} = \beta_{00k} + \beta_{01k}\text{NSF}_{jk} + r_{0jk}$

Level-3 Model: $\beta_{00k} = \gamma_{000} + u_{00k}$

We know what $\pi_{0jk}$ equals, so we can substitute that into the Level 1 model.

DV: Math score of individual student $i$

DV: Effect of class $j$

DV: Effect of school $k$
Another Look at Nested Models

Student level

Level-1 Model: \( Y_{ijk} = \beta_{00k} + \beta_{01k} NSF_{jk} + r_{0jk} \)

+ \( \pi_{1jk}(\text{Prior Achievement}_{ijk} - \bar{X}_1) \)
+ \( \pi_{2jk}(\text{Gender}_{ijk} - \bar{X}_2) \)
+ \( \pi_{3jk}(\text{Black}_{ijk} - \bar{X}_3) \)
+ \( \pi_{4jk}(\text{Hispanic}_{ijk} - \bar{X}_4) \)
+ \( \pi_{5jk}(\text{Native American}_{ijk} - \bar{X}_5) \)
+ \( \pi_{6jk}(\text{Other Ethnicity}_{ijk} - \bar{X}_6) + r_{ijk} \)

School level

Level-3 Model: \( \beta_{00k} = \gamma_{000} + u_{00k} \)

DV: Math score of individual student \( i \)

DV: Effect of school \( k \)
Another Look at Nested Models

**Student level**

Level-1 Model:  
\[ Y_{ijk} = \beta_{00k} + \beta_{01k} NSF_{jk} + r_{0jk} + \pi_{1jk} (\text{Prior Achievement}_{ijk} - \bar{X}_1) + \pi_{2jk} (\text{Gender}_{ijk} - \bar{X}_2) + \pi_{3jk} (\text{Black}_{ijk} - \bar{X}_3) + \pi_{4jk} (\text{Hispanic}_{ijk} - \bar{X}_4) + \pi_{5jk} (\text{Native American}_{ijk} - \bar{X}_5) + \pi_{6jk} (\text{Other Ethnicity}_{ijk} - \bar{X}_6) + r_{ijk} \]

*DV: Math score of individual student $i$*

Effect of being in classroom $j$ is a function of overall intercept ($\gamma_{000}$) + random school intercept $u_{00k}$

**School level**

Level-3 Model:  
\[ \beta_{00k} = \gamma_{000} + u_{00k} \]

*DV: Effect of school $k$*
Another Look at Nested Models

Level-1 Model: \( Y_{ijk} = \gamma_{000} + u_{00k} \)
+ \( \beta_{01k} \text{NSF}_{jk} + r_{0jk} \)
+ \( \pi_{1jk}(\text{Prior Achievement}_{ijk} - \bar{X}_1) \)
+ \( \pi_{2jk}(\text{Gender}_{ijk} - \bar{X}_2) \)
+ \( \pi_{3jk}(\text{Black}_{ijk} - \bar{X}_3) \)
+ \( \pi_{4jk}(\text{Hispanic}_{ijk} - \bar{X}_4) \)
+ \( \pi_{5jk}(\text{Native American}_{ijk} - \bar{X}_5) \)
+ \( \pi_{6jk}(\text{Other Ethnicity}_{ijk} - \bar{X}_6) + r_{ijk} \)

DV: Math score of individual student \( i \)
Another Look at Nested Models

Student level

Level-1 Model: \( Y_{ijk} = \gamma_{000} + u_{00k} + \beta_{01k} NSF_{jk} + r_{0jk} + \pi_{1jk}(Prior\text{\_Achievement}_{ijk} - \bar{X}_1) + \pi_{2jk}(Gender_{ijk} - \bar{X}_2) + \pi_{3jk}(Black_{ijk} - \bar{X}_3) + \pi_{4jk}(Hispanic_{ijk} - \bar{X}_4) + \pi_{5jk}(Native\text{\_American}_{ijk} - \bar{X}_5) + \pi_{6jk}(Other\text{\_Ethnicity}_{ijk} - \bar{X}_6) + r_{ijk} \)

DV: Math score of individual student \( i \)

Mixed effects model

```
lmer(MathScore ~ 1 + (1|School) + NSF + (1|Classroom) + PriorAchievement + Gender + Ethnicity, data=mathdata)
```
Week 11: Longitudinal Data & Signal Detection

- Effect Sizes & Other Follow-Up
- Longitudinal Data
  - Overview
  - Growth Curve Analysis
    - Main Effect
    - Random Slopes
    - Other Variables
    - Quadratic & Higher Degrees
- Signal Detection Theory
  - Why Do We Need SDT?
  - Sensitivity vs. Response Bias
  - Implementation
  - Interpretation & Example Analyses
  - Terminology & Theory
• Last week, we assessed each student’s math performance only once
This Week’s Design

- If we did this more than once, we’d add another level of clustering
  - Level 1 model is now the model of individual time points
• Vocab growth in the second year of life
• **Time points in children in neighborhoods**
Week 11: Longitudinal Data & Signal Detection

- Effect Sizes

- Longitudinal Data
  - Overview
  - Growth Curve Analysis
    - Main Effect
    - Random Slopes
    - Other Variables
    - Quadratic & Higher Degrees

- Signal Detection Theory
  - Why Do We Need SDT?
  - Sensitivity vs. Response Bias
  - Implementation
  - Interpretation & Example Analyses
  - Terminology & Theory
Time as a Predictor Variable

• We can add the effect of Time to our model
  • Here: Months since the study started
  • Nothing “special” about time as a predictor

• `model1 <- lmer(VocabWords ~ 1 + Time + (1|Child) + (1|Neighborhood), data=vocab)`

<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>4.0656</td>
<td>5.2226</td>
<td>0.78</td>
</tr>
<tr>
<td>Time</td>
<td>54.7802</td>
<td>0.9305</td>
<td>58.87</td>
</tr>
</tbody>
</table>

• Gain of about ~55 words per month
Time as a Predictor Variable

- **Not** necessary to have every time point represented

- Dependent variable should be on *same scale* across time points for this to be meaningful

- **Time units** don’t matter as long as they’re consistent
  - Could be hours, days, years …
**Time as a Predictor Variable**

- **Q:** What if we’re missing time points from some participants?

- **A:** Not necessary to have every time point for every participant to run the model.

- But, we can better estimate “the effect of being Subject 23” the more data we have from that participant. Estimate has greater error.

- If some participants are selectively
Time as a Predictor Variable

• **Q:** What about a repeated measures design where we have a subject complete 20 experimental trials, but we aren’t trying to look at change over time?

• **A:** Like the experimental studies we looked at early in the term. We definitely need random effects of Subject to account for the clustering of observations (repeated measures) within subjects. But, if we don’t expect any systematic changes with time, no reason to include time as a specific predictor.
Week 11: Longitudinal Data & Signal Detection

- **Effect Sizes**
- **Longitudinal Data**
  - Overview
  - Growth Curve Analysis
    - Main Effect
    - Random Slopes
    - Other Variables
    - Quadratic & Higher Degrees
- **Signal Detection Theory**
  - Why Do We Need SDT?
  - Sensitivity vs. Response Bias
  - Implementation
  - Interpretation & Example Analyses
  - Terminology & Theory
So far, we assume the same growth rate for all kids
Almost certainly not true!
At level 2, we’re sampling kids both with different starting points (intercepts) and growth rates (slopes)
Longitudinal Data: Random Slopes

RANDOM INTERCEPTS MODEL

Kids vary in starting point, but all acquire vocabulary at the same rate over this period.

WITH RANDOM SLOPES

Allows rate of vocab acquisition to vary across kids (as well as intercept)
Longitudinal Data: Random Slopes

• \texttt{model.Slope <- lmer(VocabWords ~ 1 + Time + (Time|Child) + (1|Neighborhood), data=vocab)}

\begin{verbatim}
> summary(model.Slope)
Linear mixed model fit by REML ['lmerMod']
Formula: VocabWords ~ 1 + Time + (1 + Time | Child) + (1 | Neighborhood)
Data: vocab

REML criterion at convergence: 10233.3

Scaled residuals:
    Min     1Q Median     3Q    Max
-2.5781 -0.6144 -0.1056  0.6366  3.4083

Random effects:
Groups     Name (Intercept)   Variance  Std.Dev.   Corr
Child      (Intercept)       25.241     5.024
            Time             390.777    19.768   -1.00
Neighborhood (Intercept)    7.822     2.798
                          Residual            1414.459   37.609
Number of obs: 960, groups: Child, 192; Neighborhood, 12

Fixed effects:
             Estimate Std. Error   t value
(Intercept)  7.065      2.281     3.10
Time         53.379      1.540    34.66

Correlation of Fixed Effects:
                      (Intr)     Time
(Intercept)   1.000
Time          -0.424
\end{verbatim}

In fact, LOTS of variance in Time slope
Longitudinal Data: Random Slopes

- Would also be possible to have a random slope of Time by Neighborhood
  - If there’s clustering of growth rates at the neighborhood level

```
model.TwoSlopes <- lmer(VocabWords ~ 1 + Time + (Time|Child) + (Time|Neighborhood), data=vocab)
```

- Is this any evidence for this clustering?
  - `anova(model.Slope, model.TwoSlopes)`

<table>
<thead>
<tr>
<th>Df</th>
<th>AIC</th>
<th>BIC</th>
<th>loglik</th>
<th>deviance</th>
<th>Chi^2</th>
<th>Chi DF</th>
<th>Pr(&gt;Chi^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>10253</td>
<td>10287</td>
<td>-5119.6</td>
<td>10239</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>10257</td>
<td>10300</td>
<td>-5119.3</td>
<td>10239</td>
<td>0.6114</td>
<td>2</td>
<td>0.7366</td>
</tr>
</tbody>
</table>

χ^2(2) = 0.61
p = .74
n.s.
Week 11: Longitudinal Data & Signal Detection

- Effect Sizes
  - Longitudinal Data
    - Overview
    - Growth Curve Analysis
      - Main Effect
      - Random Slopes
      - Other Variables
      - Quadratic & Higher Degrees
  - Signal Detection Theory
    - Why Do We Need SDT?
    - Sensitivity vs. Response Bias
    - Implementation
    - Interpretation & Example Analyses
    - Terminology & Theory
Other Variables

- We may want to include other variables in a longitudinal model:
  - Do parents frequently read picture books to the child?

- **Time + Reading**
  - Effect of Reading *invariant across time*
  - Can only affect the intercept (parallel lines)

- **Time * Reading**
  - Effect of Reading *varies with time*
  - Can affect intercept & slope
**Other Variables: Results**

- Model results with the 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>5.020</td>
<td>3.120</td>
<td>1.609</td>
</tr>
<tr>
<td>Time</td>
<td>50.040</td>
<td>2.146</td>
<td>23.313</td>
</tr>
<tr>
<td>ReadingYes</td>
<td>3.908</td>
<td>4.230</td>
<td>0.924</td>
</tr>
<tr>
<td>Time:ReadingYes</td>
<td>6.703</td>
<td>3.049</td>
<td>2.198</td>
</tr>
</tbody>
</table>

Growth rate for “No” group:
50.040 words / month

Growth rate for “Yes” group:
50.040 + 6.703 = 56.743 words / month

*Parental reading doesn’t affect vocab at time 0*

*But, results in faster vocab growth (amplifies + Time effect)*

e.g., Huttenlocher et al., 1991
Week 11: Longitudinal Data & Signal Detection

- Effect Sizes
  - Longitudinal Data
    - Overview
    - Growth Curve Analysis
      - Main Effect
      - Random Slopes
      - Other Variables
      - Quadratic & Higher Degrees
  - Signal Detection Theory
    - Why Do We Need SDT?
    - Sensitivity vs. Response Bias
    - Implementation
    - Interpretation & Example Analyses
    - Terminology & Theory
Growth Curve Analysis

• We’ve been assuming a *linear* effect of time

• But, it looks like vocab growth may *accelerate*
  • Growth between 2 mo. and 4 mo. is much smaller than growth between 6 mo. and 8 mo.
  • Suggests a *curve / quadratic equation*
Growth Curve Analysis

• Add quadratic effect ($\text{Time}^2$):
  • `model.poly <- lmer(VocabWords ~ 1 + poly(Time, degree=2, raw=TRUE) + (1 + poly(Time, degree=2, raw=TRUE) | Child) + (1 | Neighborhood), data=vocab)`

  • `degree=2` because we want $\text{Time}^2$

  • `poly()` automatically adds lower-order terms as well
    • i.e., the linear term ($\text{Time}$)
Growth Curve Analysis: Results

• 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>39.3533</td>
<td>2.4170</td>
<td>16.28</td>
</tr>
<tr>
<td>poly(Time, degree = 2, raw = TRUE)1</td>
<td>10.6299</td>
<td>0.6677</td>
<td>15.92</td>
</tr>
<tr>
<td>poly(Time, degree = 2, raw = TRUE)2</td>
<td>6.9948</td>
<td>0.2145</td>
<td>32.61</td>
</tr>
</tbody>
</table>

• Implied equation (approximate):
  • \( \text{VocabWords} = 40 + 11 \times \text{Time} + 7 \times \text{Time}^2 \)

• Some predicted values:
  • If time=0, \( \text{VocabWords} = 40 + (11 \times 0) + (7 \times 0^2) = 40 \)
  • If time=1, \( \text{VocabWords} = 40 + (11 \times 1) + (7 \times 1^2) = 58 \)
  • If time=2, \( \text{VocabWords} = 40 + (11 \times 2) + (7 \times 2^2) = 90 \)
  • Vocab growth is accelerating (larger change from time 1 to time 2 than from time 0 to time 1)
Growth Curve Analysis

• Could go up to even higher degrees (Time$^3$, Time$^4$…)
  • \textit{degree}=3 if highest exponent is 3

• Degree minus 1 = Number of \textit{bends} in the curve
Growth Curve Analysis

- **Maximum** degree of polynomial: # of time points minus 1
  - Example: 2 time points perfectly fit by a line (degree 1). Nothing left for a quadratic term to explain.

- But, don’t want to **overfit**
  - Probably not the case that the real underlying (population) trajectory has 6 bends in it

- What degree **should** we include?
  - Theoretical considerations
  - If comparing conditions, look at mean trajectory across conditions (Mirman et al., 2008)
Week 11: Longitudinal Data & Signal Detection

- Effect Sizes
- Longitudinal Data
  - Overview
  - Growth Curve Analysis
    - Main Effect
    - Random Slopes
    - Other Variables
    - Quadratic & Higher Degrees
- Signal Detection Theory
  - Why Do We Need SDT?
  - Sensitivity vs. Response Bias
  - Implementation
  - Interpretation & Example Analyses
  - Terminology & Theory
Tasks With Categorical Decisions

las gatos

(1) Grammatical
(4) Ungrammatical

The cop saw the spy with the binoculars.

In analyzing these decisions, need to consider both overall preference for certain categories & judgments of individual items.
Study:

POTATO
SLEEP
RACCOON
WITCH
NAPKIN
BINDER
• Test:
• SLEEP
• POTATO
• BINDER
• WITCH
• RACCOON
• NAPKIN
In early memory experiments, all test probes were previously studied items. No way to distinguish a person who actually remembers everything from a person who’s realized these are ALL “old” items.
Adding “lure” items helps make the task less obvious. But still have to interpret response to lures. Did this person circle 50% of studied items because they remember seeing those words … or because they circled 50% of everything?
Signal Detection Theory

- For analyzing **categorical judgments**
  - Part **method for analyzing** judgments
  - Part **theory** about how people **make judgments**

- Originally developed for psychophysics

- Purpose:
  - Better metric properties than ANOVA on proportions
    (logistic regression has already taken care of this)
  - Distinguish **sensitivity** from **response bias**
Week 11: Longitudinal Data & Signal Detection

- Effect Sizes
- Longitudinal Data
  - Overview
  - Growth Curve Analysis
    - Main Effect
    - Random Slopes
    - Other Variables
    - Quadratic & Higher Degrees
- Signal Detection Theory
  - Why Do We Need SDT?
  - Sensitivity vs. Response Bias
  - Implementation
  - Interpretation & Example Analyses
  - Terminology & Theory
Sensitivity vs. Response Bias

“If you’re not sure, guess C”

Response bias

Knowing which answers are C and which aren't

Sensitivity
Sensitivity vs. Response Bias

- Imagine asking groups of second-language learners of English to judge grammaticality...

<table>
<thead>
<tr>
<th>Without Intervention</th>
<th>ACCURACY</th>
<th>SAID “GRAMMATICAL”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grammatical condition</td>
<td>80%</td>
<td>80%</td>
</tr>
<tr>
<td>Ungrammatical cond.</td>
<td>20%</td>
<td>80%</td>
</tr>
</tbody>
</table>

People just judge 80% of sentences grammatical in both conditions.

This is all response bias—no evidence that they are sensitive to whether particular sentences are grammatical or not.
Similarly, an intervention could shift response bias without actually increasing sensitivity.

<table>
<thead>
<tr>
<th>Without Intervention</th>
<th>ACCURACY</th>
<th>SAID &quot;GRAMMATICAL&quot;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grammatical condition</td>
<td>80%</td>
<td>80%</td>
</tr>
<tr>
<td>Ungrammatical cond.</td>
<td>20%</td>
<td>80%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>With Intervention</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Grammatical condition</td>
<td>60%</td>
<td>60%</td>
</tr>
<tr>
<td>Ungrammatical cond.</td>
<td>40%</td>
<td>60%</td>
</tr>
</tbody>
</table>
Sensitivity vs. Response Bias: Examples

• We present radiologists with 20 X-rays. Half of the X-rays show lung disease and half show healthy lungs. For each X-ray, the radiologist has to judge whether lung disease is present.

• In this study, how can we define…
  • Response bias?
  • Sensitivity?
Sensitivity vs. Response Bias: Examples

- We present radiologists with 20 X-rays. Half of the X-rays show lung disease and half show healthy lungs. For each X-ray, the radiologist has to judge whether lung disease is present.

- In this study, how can we define…
  - Response bias?
    - Overall propensity to judge that lung disease is present
  - Sensitivity?
    - Does the radiologist diagnose the patient with lung disease more in the cases where the patient actually has lung disease?
Sensitivity vs. Response Bias: Examples

• We present undergraduates with a series of moral dilemmas in which they have to imagine deciding between saving 1 person’s life and saving several people’s lives. The dependent measure is how often people make the utilitarian choice to save several people. Some scenarios are less personal, and we hypothesize that people will make more utilitarian choices in these scenarios.

• In this study, how can we define…
  • Response bias?

  • Sensitivity?
Sensitivity vs. Response Bias: Examples

• We present undergraduates with a series of moral dilemmas in which they have to imagine deciding between saving 1 person’s life and saving several people’s lives. The dependent measure is how often people make the utilitarian choice to save several people. Some scenarios are less personal, and we hypothesize that people will make more utilitarian choices in these scenarios.

• In this study, how can we define…
  • Response bias?
    • Overall frequency of utilitarian judgments
  • Sensitivity?
    • Do people make more of the utilitarian judgments when the scenario is less personal?
**Sensitivity vs. Response Bias: Examples**

- We ask college students studying French to proofread a set of 40 French sentences, all of which contain a subject/verb agreement error. The dependent measure is whether or not the student judge the sentence as containing a subject/verb agreement error (i.e., “error” or “no error”).

- In this study, how can we define…
  - Response bias?
  - Sensitivity?
Sensitivity vs. Response Bias: Examples

• We ask college students studying French to proofread a set of 40 French sentences, all of which contain a subject/verb agreement error. The dependent measure is whether or not the student judge the sentence as containing a subject/verb agreement error (i.e., “error” or “no error”).

• In this study, how can we define…
  • Response bias?

• Sensitivity?

Trick question!! This is like the memory test that contains only “old” items. Because the test *only* contains errors, there’s no way to tell whether a participant’s response is driven by their general bias to report errors or by noticing the error in this specific sentence. We cannot separate response bias from sensitivity here. Unfortunately, this limits the conclusions we can draw from this task.
Sensitivity vs. Response Bias

- Comparison to “chance” get at a similar idea
  - But, such comparisons assumes all responses equally likely

- Many experiments do balance frequency of intended responses

- But even so, bias can differ for many reasons
  - Relative frequency in experiment
  - Prior frequency in the world (“no disease” less common than “disease”)
  - Motivational factors (e.g., one error “less bad” than another)
Week 11: Longitudinal Data & Signal Detection

- Effect Sizes
- Longitudinal Data
  - Overview
  - Growth Curve Analysis
    - Main Effect
    - Random Slopes
  - Other Variables
  - Quadratic & Higher Degrees
- Signal Detection Theory
  - Why Do We Need SDT?
  - Sensitivity vs. Response Bias
    - Implementation
    - Interpretation & Example Analyses
    - Terminology & Theory
Example Study: Fraundorf, Watson, & Benjamin (2010)

Both the British and the French biologists had been searching Malaysia and Indonesia for the endangered monkeys.

Finally, the **British** spotted one of the monkeys in Malaysia and planted a radio tag on it.

Emphasized or not?
The British scientists spotted the endangered monkey and tagged it.
The French scientists spotted the endangered monkey and tagged it.
SDT & Mixed Effects Models

- Traditional logistic regression model:
  
  \[ \text{Accuracy of Response} = \text{Probe Type} \times \text{Emphasis} \]

- **Accuracy** confounds **sensitivity** and **response bias**
  - Manipulation might just make you say *true* to everything more
**SDT & Mixed Effects Models**

- Traditional logistic regression model:
  \[ \text{Accuracy of Response} = \text{Probe Type} \times \text{Emphasis} \]
  - Correct Memory or Incorrect Memory

- Signal detection model:
  \[ \text{Response Made} = \text{Probe Type} \times \text{Emphasis} \]
  - Judged “True” vs Judged “False”
  - Judged “Grammatical” or Judged “Ungrammatical”
Respond correctly or Respond incorrectly?

True statement or False statement?
Week 11: Longitudinal Data & Signal Detection

- Effect Sizes
- Longitudinal Data
  - Overview
  - Growth Curve Analysis
    - Main Effect
    - Random Slopes
    - Other Variables
    - Quadratic & Higher Degrees
- Signal Detection Theory
  - Why Do We Need SDT?
  - Sensitivity vs. Response Bias
  - Implementation
  - Interpretation & Example Analyses
  - Terminology & Theory
**SDT & Mixed Effects Models**

- **SDT model:**

  - **Said “TRUE”**
  - **Intercept**
    - Baseline rate of responding TRUE.
  - **Probe Type is TRUE**
    - Does item being true make you more likely to say TRUE?
  - **Contrastive Emphasis**
    - Does contrastive emphasis change overall rate of saying TRUE?
  - **Accent x TRUE**
    - Does accent especially increase TRUE responses to true items?

  Overall response bias

  Overall sensitivity

  Effect on bias

  Effect on sensitivity

With centered predictors...
SDT & Mixed Effects Models

- SDT model:

  - Said “TRUE”
    - w/ centered predictors...
    - Baseline rate of responding TRUE.

  - Intercept
    - Does item being true make you more likely to say TRUE?

  - Probe Type is TRUE
    - Does contrastive emphasis change overall rate of saying TRUE?

  - Contrastive Emphasis
    - Effect on bias

  - Accent x TRUE
    - Does accent especially increase TRUE responses to true items?

Overall response bias
Overall sensitivity
Effect on bias
Effect on sensitivity

• When & how do people avoid ambiguity in what they say?

• Task: Read sentences & repeat back from memory

• **Ambiguous** sentence start: “The coach knew you…”
  – “The coach knew you since freshman year.” *(knowing you)*
  – “The coach knew you missed practice.” *(knowing a fact)*

• “The coach knew *that* you…”
  • “*that*” is optional but clarifies it’s a *knowing-a-fact* sentence
  • Dependent measure: Do people say “*that*” here?

• Are people **sensitive** to diff. from unambiguous case?:
  • “The coach knew *I*…”
    • *Knowing-a-person* sentence would be “The coach knew *me*.”

• Also vary whether instructions emphasize being clear
SDT & Multi-Level Models

- SDT model:
  
  \[
  \text{Said "that"} = \text{Intercept} + \text{Ambiguity} + \text{Instructions} + \text{Instructions x Ambiguity} \\
  \]

  Baseline rate of including "that" w/ centered predictors...

  Overall response bias

  Do people say "that" more for you (unambig.) than for I (ambig.)

  Overall sensitivity

  Are people told to avoid ambiguity?

  Effect on bias

  Do instructions especially increase use of "that" for ambiguous items?

  Effect on sensitivity
**SDT & Multi-Level Models**

- **SDT model:**

  Said “that”

  \[ \text{Baseline rate of including “that”} \]

  w/ centered predictors...

  \[ \text{Overall response bias} \]

  \[ \text{Overall sensitivity} \]

  \[ \text{Effect on sensitivity} \]

  \[ \text{Effect on bias} \]

**Results**

- **Intercept**

  Do people say “that” more for you (unambig.) than for I (ambig.)

  \[ \text{Effect on sensitivity} \]

- **Instructions**

  Are people told to avoid ambiguity?

  \[ \text{Effect on bias} \]

- **Instructions x Ambiguity**

  Do instructions especially increase use of “that” for ambiguous items?

  \[ \text{Effect on sensitivity} \]

- People NOT sensitive to whether what they’re saying is grammatically ambiguous

- Effect of emphasizing clarity is that people just add extra “that”s everywhere (whether actually needed or not)
  - Case where a change in response bias tells us something interesting about what people are doing

- Response bias is NOT just something we want to avoid / get rid of
  - Can be theoretically interesting

- Our measure of sensitivity in the SDT model is independent of response bias, so OK to look at sensitivity even if there is a response bias effect
Week 11: Longitudinal Data & Signal Detection

- Effect Sizes
- Longitudinal Data
  - Overview
  - Growth Curve Analysis
    - Main Effect
    - Random Slopes
    - Other Variables
    - Quadratic & Higher Degrees
- Signal Detection Theory
  - Why Do We Need SDT?
  - Sensitivity vs. Response Bias
  - Implementation
  - Interpretation & Example Analyses
  - Terminology & Theory – courtesy Jason Finley
# Signal Detection Performance

## True State

<table>
<thead>
<tr>
<th>True State</th>
<th>Signal</th>
<th>Noise</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signal</td>
<td>Hit</td>
<td>False Alarm</td>
</tr>
<tr>
<td>Noise</td>
<td>Miss</td>
<td>Correct Rejection</td>
</tr>
</tbody>
</table>

## Observer's Response

- **“Yes”**
- **“No”**

**Summary Statistics:**

- Hit Rate (HR) = \( \frac{\text{# Hits}}{\text{# Signal Trials}} \)
- False Alarm Rate (FAR) = \( \frac{\text{# False Alarms}}{\text{# Noise Trials}} \)
1. Trials = events
2. Strength of evidence: continuous dimension
3. Conditional probability distributions for noise, signal
4. Decision/response criterion
Response Bias

Shifts in response bias (response criterion) change relative probability of these categories.
Response Bias

Higher criterion = more correct rejections, but also more misses.

<table>
<thead>
<tr>
<th>True State</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Signal</td>
<td></td>
</tr>
<tr>
<td>Noise</td>
<td></td>
</tr>
<tr>
<td>Hit</td>
<td></td>
</tr>
<tr>
<td>False Alarm</td>
<td></td>
</tr>
<tr>
<td>Miss</td>
<td></td>
</tr>
<tr>
<td>Correct Rejection</td>
<td></td>
</tr>
</tbody>
</table>

![Response Bias Diagram](image)

- **NOISE trials**
- **SIGNAL trials**

- **Probability**
- **Evidence**

- **decision criterion**
Response Bias

Shifts in response bias (response criterion) change relative probability of these categories.

Response Bias

<table>
<thead>
<tr>
<th>True State</th>
<th>Signal</th>
<th>Noise</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;Yes&quot;</td>
<td>Hit</td>
<td>False Alarm</td>
</tr>
<tr>
<td>&quot;No&quot;</td>
<td>Miss</td>
<td>Correct Rejection</td>
</tr>
</tbody>
</table>

![Graph of Probability vs. Evidence](image)

Probability

Evidence

Decision criterion

NOISE trials

SIGNAL trials
Response Bias

Lower criterion = more hits, but also more false alarms.

c = \(-.5[z(\text{HR}) + z(\text{FAR})]\)

<table>
<thead>
<tr>
<th>True State</th>
<th>Signal</th>
<th>Noise</th>
</tr>
</thead>
<tbody>
<tr>
<td>Response</td>
<td>“Yes”</td>
<td>“No”</td>
</tr>
<tr>
<td>Hit</td>
<td>Hit</td>
<td>Miss</td>
</tr>
<tr>
<td>False Alarm</td>
<td>False Alarm</td>
<td>Correct Rejection</td>
</tr>
</tbody>
</table>
Sensitivity

\[ d' = z(\text{HR}) - z(\text{FAR}) \]
Sensitivity

Lower sensitivity is equivalent to *more similar (less distinct) distributions of evidence*. 

![Graph showing Lower Sensitivity](image_url)
Sensitivity

Higher sensitivity is equivalent to less similar (more distinct) distributions of evidence.
Effect Sizes

Longitudinal Data
  - Overview
  - Growth Curve Analysis
    - Main Effect
    - Random Slopes
    - Other Variables
    - Quadratic & Higher Degrees

Signal Detection Theory
  - Why Do We Need SDT?
  - Sensitivity vs. Response Bias
  - Implementation
  - Interpretation & Example Analyses
  - Terminology & Theory
Extensions

- In principle, generalizes > 2 categories:
  - Collapse over items & use $d_a$
  - Mixed effects version with crossed random effects would need ordinal model—not yet available
Extensions

- Unequal variance
  - So far, variability of response is a constant
    \[ d' = B_0 + B_1X_1 + (1|Subject) + \varepsilon \]
  - Definitely *not* true for recognition memory

- Noisy criterion (Benjamin, Diaz, & Wee, 2009)
  - Typically, all error is in the *response* (evidence)
  - But criterion could vary from trial to trial