Week 13.1: Signal Detection Theory

- Signal Detection Theory
- Why Do We Need SDT?
  - Sensitivity vs. Response Bias
  - Implementation
  - SDT & Other Independent Variables
  - Logit vs. Probit
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.
Study:

POTATO
SLEEP
RACCOON
WITCH
NAPKIN
BINDER

Test:

• SLEEP
• POTATO
• HEDGE
• BINDER
• SHELL
• RACCOON
• MONKEY
• OATH

• 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 13.1: Signal Detection Theory

- Signal Detection Theory
- Why Do We Need SDT?
  - Sensitivity vs. Response Bias
    - Implementation
    - SDT & Other Independent Variables
    - Logit vs. Probit
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...
**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></th>
<th>Without Intervention</th>
<th></th>
<th>With Intervention</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ACCURACY</td>
<td>SAID</td>
<td>“GRAMMATICAL”</td>
</tr>
<tr>
<td>Grammatical condition</td>
<td>80%</td>
<td>80%</td>
<td>60%</td>
</tr>
<tr>
<td>Ungrammatical cond.</td>
<td>20%</td>
<td>80%</td>
<td>60%</td>
</tr>
<tr>
<td>Grammatical condition</td>
<td>60%</td>
<td></td>
<td>60%</td>
</tr>
<tr>
<td>Ungrammatical cond.</td>
<td>40%</td>
<td></td>
<td>60%</td>
</tr>
</tbody>
</table>
Sensitivity vs. Response Bias

- Proportion accuracy would be misleading
- We want an analysis that tests both subjects’ sensitivity and their response bias

<table>
<thead>
<tr>
<th></th>
<th>ACCURACY</th>
<th>SAID “GRAMMATICAL”</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Without Intervention</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grammatical condition</td>
<td>80%</td>
<td>80%</td>
</tr>
<tr>
<td>Ungrammatical cond.</td>
<td>20%</td>
<td>80%</td>
</tr>
<tr>
<td><strong>With Intervention</strong></td>
<td></td>
<td></td>
</tr>
<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>
Comparison to “chance” get at a similar idea
  But, that 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)
– Not bad to have a response bias—we just need to account for it in our analysis!
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 are conducting a cross-cultural study of color perception. Participant in a variety of cultures each see 40 pairs of paint chips. For every pair, the participant judges if the two chips are the same color or different colors. In reality, 20 pairs are the same color, and 20 pairs are different colors.

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

• We are conducting a cross-cultural study of color perception. Participant in a variety of cultures each see 40 pairs of paint chips. For every pair, the participant judges if the two chips are the same color or different colors. In reality, 20 pairs are the same color, and 20 pairs are different colors.
• In this study, how can we define…
  • Response bias?
    • *Overall* tendency to judge pairs as the same
  • Sensitivity?
    • Do people judge pairs as the same more when they are *actually* the same?
Sensitivity vs. Response Bias: Examples

• We are conducting a cross-cultural study of color perception. Participant in a variety of cultures each see 40 pairs of paint chips. For every pair, the participant judges if the two chips are the same color or different colors. In reality, 20 pairs are the same color, and 20 pairs are different colors.
• In this study, how can we define…
  • Response bias?
    • Overall tendency to judge pairs as the same
  • Sensitivity?
    • Do people judge pairs as the same more when they are actually the same?
Sensitivity vs. Response Bias: Examples

• An I/O psychologist is interested in how extracurricular activities influence the post-college job search. Each research participant sees a series of fictitious resumes and, for each resume, judges whether they think the person merits hiring. The researcher experimentally varies the number of extracurricular activities listed on the resumes.

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

  • Sensitivity?
Sensitivity vs. Response Bias: Examples

• An I/O psychologist is interested in how extracurricular activities influence the post-college job search. Each research participant sees a series of fictitious resumes and, for each resume, judges whether they think the person merits hiring. The researcher experimentally varies the number of extracurricular activities listed on the resumes.

• In this study, how can we define…
  • Response bias?
    • Overall tendency to think people merit hiring
  • Sensitivity?
    • Do extracurricular activities increase hiring?
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.
Week 13.1: Signal Detection Theory

- Signal Detection Theory
- Why Do We Need SDT?
- Sensitivity vs. Response Bias
- Implementation
  - SDT & Other Independent Variables
  - Logit vs. Probit
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.
The British scientists spotted the endangered monkey and tagged it.

Probe type = TRUE
The French scientists spotted the endangered monkey and tagged it.
SDT & Mixed Effects Models

- Traditional logistic regression model:
  \[
  \text{Correct} \sim 1 + \text{ProbeType}
  \]

Accuracy confounds sensitivity and response bias
  - Accuracy might differ across probe types just because of bias to respond true
SDT & Mixed Effects Models

- Traditional logistic regression model:

  Correct ~ 1 + ProbeType

- Signal detection model:

  JudgmentMade ~ 1 + ProbeType
SDT & Mixed Effects Models

- Traditional logistic regression model:

  Correct ~ 1 + ProbeType

  CORRECT MEMORY or INCORRECT MEMORY?

- More generally:

  \[
  \text{glmer(JudgmentMade} \sim 1 + \text{StimulusCategory} + (1|\text{RandomEffect}), \text{data=} \text{dataname, family=} \text{binomial})
  \]
Respond correctly or respond incorrectly?

True statement or False statement?
**SDT & Mixed Effects Models**

- SDT model:
  
  Said “TRUE” = Intercept + Probe Type is TRUE

  \[ \text{JudgmentMade} \sim 1 + \text{ProbeType} \]

  *w/ effects coding…*

  Overall response bias

  Baseline rate of responding TRUE.

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

  Sensitivity
SDT & Mixed Effects Models

- SDT model:

  JudgmentMade ~ 1 + ProbeType

  w/ effects coding...

  Baseline rate of responding TRUE.

  Overall response bias

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

  Sensitivity
**SDT & Mixed Effects Models**

- More generally:

  \[
  \text{JudgmentMade} \sim 1 + \text{StimulusType}
  \]

  w/ effects coding...

  **Intercept**
  
  Baseline rate of “A” responses

  **Stimulus Type**
  
  Does item *being* in category “A” make you more likely to judge as ”A”?

  **Overall response bias**

  **Sensitivity**
Now You Try It!

- bpd.csv
- Clinical trainees evaluating learning to diagnose borderline personal disorder (BPD). Each trainee sees 60 cases—half with BPD and half without—and makes a diagnosis for each.

- Potentially relevant columns:
  - **JudgedBPD**: Trainees’ judgment of BPD (1 yes, 0 no)
  - **HasBPD**: Whether the person in the case actually has BPD—as diagnosed by expert (“Y” or “N”)
  - **Accuracy**: Was the trainees’ judgment correct? (1 yes, 0 no)
Now You Try It!

- If our memory experiment SDT analysis involved a model formula like this:

\[ \text{JudgmentMade} \sim 1 + \text{ProbeType} + (1|\text{Subject}) \]

- Can you run a SDT model on the \textit{bpd} data?
  - \textit{Tip 1:} Apply effects coding (-0.5 and 0.5) to the predictor variable!
  - \textit{Tip 2:} Should this be an \textit{lmer} model or a \textit{glmer} model?
Now You Try It!

- If our memory experiment SDT analysis involved a model formula like this:
  
  JudgmentMade ~ 1 + ProbeType + (1|Subject)

- Can you run a SDT model on the bpd data?
  - contrasts(bpd$HasBPD) <- c(-0.5, 0.5)
  - model1 <- glmer(JudgedBPD ~ 1 + HasBPD + (1|Trainee), family=binomimal, data=bpd)
### Now You Try It!

**Random effects:**

<table>
<thead>
<tr>
<th>Groups</th>
<th>Name</th>
<th>Variance</th>
<th>Std.Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trainee</td>
<td>(Intercept)</td>
<td>0.07657</td>
<td>0.2767</td>
</tr>
</tbody>
</table>

**Number of obs:** 3600, **groups:** Trainee, 60

**Fixed effects:**

| Estimate | Std. Error | z value | Pr(>|z|)  |
|----------|------------|---------|----------|
| (Intercept) | 0.04902     | 0.04943 | 0.992    | 0.321    |
| HasBPD1    | 0.71395     | 0.06844 | 10.431   | <2e-16 *** |

**Intercept:** Overall tendency to judge people as having BPD or not
- Response bias (here, not significant)

**HasBPD:** Do we get more “has BPD” judgments when the person actually has BPD?
- Sensitivity (significant!)
Now You Try It!

Our model of the random effects is that trainees differ only in their intercept

- They differ only in response bias … *not* in sensitivity

Can we also allow the sensitivity to be *different for each* trainee?
Now You Try It!

```
model2 <- glmer(JudgedBPD ~ 1 + HasBPD + (1 + HasBPD|Trainee),
                 family=binomial, data=bpd)
```
Week 13.1: Signal Detection Theory

- Signal Detection Theory
  - Why Do We Need SDT?
  - Sensitivity vs. Response Bias
- Implementation
- SDT & Other Independent Variables
  - Logit vs. Probit
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.

We now have an additional independent variable.
Signal detection model with another independent variable:

```r
my.model <- glmer(
  JudgmentMade ~ 1 + ProbeType*Emphasis
  + (1|Trainee),
  family=binomial,
  data=memory)
```

JUDGED “TRUE” OR JUDGED “FALSE”
SDT & Other Independent Variables

- More generally...

```r
my.model <- glmer(JudgmentMade ~ 1 + StimulusType*OtherIV + (1|RandomEffect),
family=binomial,
data=mydata)
```
**SDT & Other Independent Variables**

- **SDT model:**

  \[ \text{Said “TRUE”} = \text{Intercept} + \text{Probe Type is TRUE} + \text{Contrastive Emphasis} + \text{Emphasis x TRUE} \]
  
  w/ effects coding...

  - **Baseline rate of responding TRUE.** (Overall response bias)
  - **Does item \textit{being} true make you more likely to say TRUE?** (Overall sensitivity)
  - **Does contrastive emphasis change overall rate of saying TRUE?** (Effect on bias)
  - **Does emphasis especially increase TRUE responses to true items?** (Effect on sensitivity)
SDT & Other Independent Variables

- SDT model:

  **Said “TRUE”**

  w/ effects coding…

  **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?

  **Emphasis x TRUE**
  
  Does emphasis especially increase TRUE responses to true items?
SDT & Other Independent Variables

More generally...

- **Responded w/ category A**
  - w/ effects coding...
  - **Intercept**: Baseline rate of “A” responses

- **Stimulus Type**
  - Does item *being* in category “A” make you more likely to judge as “A”?

- **OtherIV**
  - Does other independent variable change overall rate of saying “A”?

- **Interaction**
  - Does other IV increase ability to identify which category item is in?

• 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 sophomore 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 & Other Independent Variables

- SDT model:

\[ \text{Said "that"} = \text{Intercept} + \text{Ambiguity} + \text{Instructions} + \text{Instructions} \times \text{Ambiguity} \]

- Baseline rate of including "that"
- Do people say "that" more for you (unambig.) than for I (ambig.)
- Are people told to avoid ambiguity?
- Do instructions especially increase use of "that" for ambiguous items?

Overall response bias
Overall sensitivity
Effect on bias
Effect on sensitivity
w/ effects coding...
**SDT & Other Independent Variables**

- **SDT model:**

  - Said “that”

  =

  - Intercept: Baseline rate of including “that”

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

  - Instructions: Are people told to avoid ambiguity?

  - Instructions x Ambiguity: Do instructions especially increase use of “that” for ambiguous items?

  **Overall response bias**

  **Overall sensitivity**

  **Effect on bias**

  **Effect on sensitivity**

w/ effects coding

- 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
Back to Our BPD Data…

- We’re concerned that there may be a gender bias in diagnoses of BPD (e.g., Bjorklund, 2009; Skodol & Bender, 2003)

- Can you test whether gender affects response bias and/or sensitivity in your model?
  - Don’t forget to apply effects coding (-0.5 and 0.5) to gender
  - Which gender do we think will get more BPD diagnoses?
Back to Our BPD Data…

- We’re concerned that there may be a Gender bias in diagnoses of BPD (e.g., Bjorklund, 2009; Skodol & Bender, 2003)

- Can you test whether Gender affects response bias and/or sensitivity in your model?
  - contrasts(bpd$Gender) <- c(0.5, -0.5)
  - model3 <- glmer(JudgedBPD ~ 1 + HasBPD*Gender + (1+HasBPD*Gender|Trainee), family=binomial, data=bpd)
Back to Our BPD Data…

Intercept: Overall tendency to judge people as having BPD or not
  - Response bias (here, not significant)
HasBPD: Do we get more “has BPD” judgments when the person actually has BPD?
  - Sensitivity (significant!)
Gender: An effect of BPD on “has BPD” judgments, regardless of whether the person has BPD
  - This an effect of gender on response bias!
Gender:HasBPD: Is “has BPD” larger for one gender?
  - No – no effect of gender on sensitivity

| Fixed effects:          | Estimate | Std. Error | z value | Pr(>|z|) |
|-------------------------|----------|------------|---------|----------|
| (Intercept)             | 0.0472   | 0.0519     | 0.910   | 0.363    |
| HasBPD1                 | 0.7449   | 0.0713     | 10.435  | <2e-16   *** |
| Gender1                 | 0.7948   | 0.0783     | 10.154  | <2e-16   *** |
| HasBPD1:Gender1         | -0.0009  | 0.1459     | -0.007  | 0.995    |
Summary:

- No overall response bias to judge people as having BPD or not
- Trainees have some ability to discern which people have BPD and which don’t
- Overall bias to diagnosis more women with BPD, but doesn’t affect sensitivity to the symptoms in making the diagnosis
Week 13.1: Signal Detection Theory

- Signal Detection Theory
  - Why Do We Need SDT?
  - Sensitivity vs. Response Bias
  - Implementation
  - SDT & Other Independent Variables
- Logit vs. Probit
Logit and Probit

• How to link the binomial response to the continuous model predictors?

• So far, we’ve been using the logit:
  \[ \text{logit} = \log \left( \frac{p(\text{recall})}{1-p(\text{recall})} \right) \]

• Probit: Based on the cumulative distribution function of the normal
  \[ d' = \text{CDF}(\text{recall}) - \text{CDF}(1-\text{recall}) \]
Logit and Probit

- Extremely similar, but logit a little less sensitive to extreme values
  - Thus, will probably get qualitatively the same results
- Which to choose?
  - Some literatures (SDT) use $d'$ units -> **Probit** model
  - Otherwise, **logit** has a somewhat easier interpretation
    - Odds / odds ratios
**Probit**

- To use the probit instead of the logit:
  - `model.Probit <- glmer(JudgedBPD ~ 1 + HasBPD + (1 + HasBPD|Trainee), data=bpd, family=binomial(link='probit'))`

- `(link='logit')` is the same as the default model