Lecture 1: Robust Cognitive Models

Jeff Rouder

Dec. 2021

Robust Modeling in Cognitive Science

Confession: I don't know what to teach.

Robust Modeling in Cognitive Science

Robust Modeling in Cognitive Science

- I don't want to repeat open science

Robust Modeling in Congnitive Science

- How about, *Everything you wanted to know about modeling but were afraid to ask?*
- EYWAKAMBWATA

EYWAKAMBWATA?

- Philosophy behind modeling
- Model Notation
- Constraint, Flexibility, Tertiary Assumptions
- Judicious Specification
- Writing Down Likelihoods
- Analysis, Comparison, Checking
- Ethics?

Motivating Problem

Is unequal-variance signal-detection a good model for recognition memory or perceptual decision making?

Signal Detection Experiment

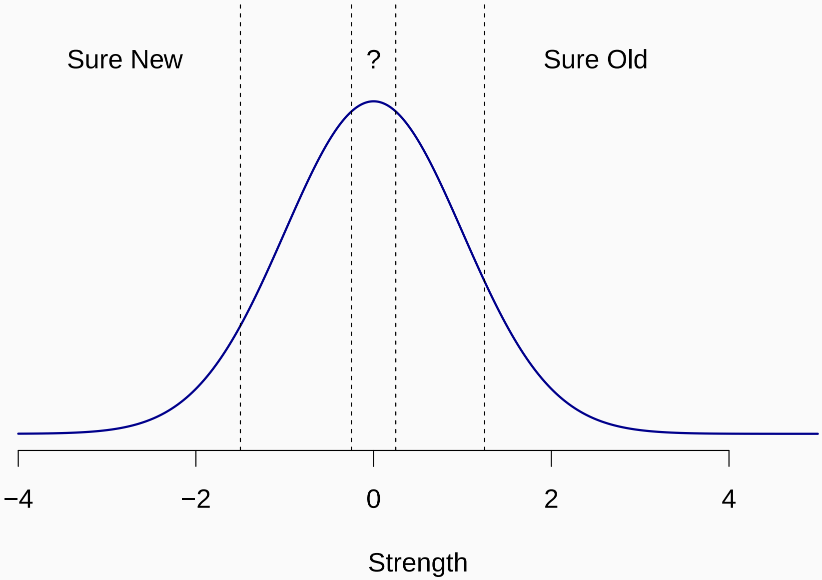
- Study: a bunch of items
- Test: some old, some new
- Judgment: Sure old, maybe old, not sure, maybe new, sure new

Signal Detection Experiment

	Sure New	Maybe New	Unsure	Maybe Old	Sure Old
Old Items	4	12	15	29	40
New Items	9	27	18	36	10

Signal Detection Model

Signal Detection Model



Your Turn

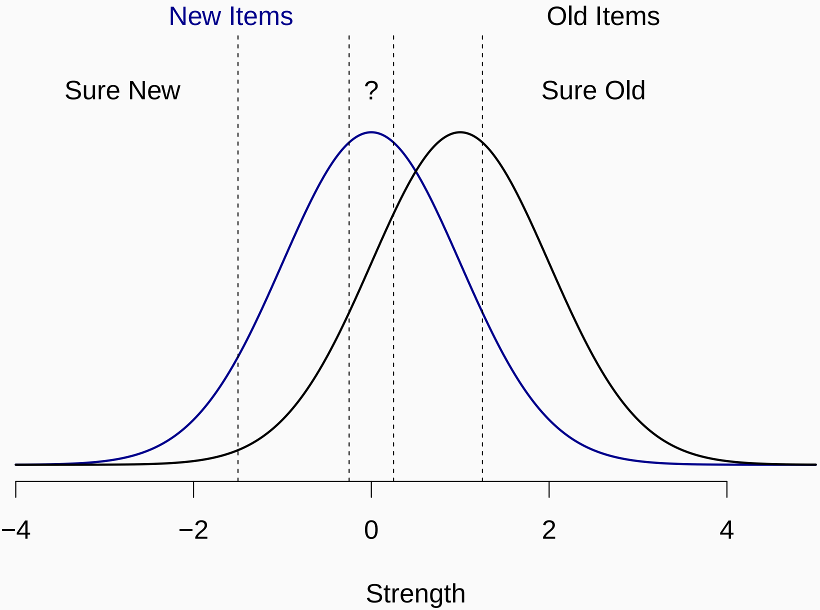
- Using `R`, code up a simulation model for new items.
- Use your simulation code to predict the probability of any response for any set of four bounds?

My Turn

```r
M=1000
x=rnorm(M)
bound=seq(-1,1,length=4)
regionArray=outer(x,bound,">")
region=apply(regionArray,1,sum)+1
table(region)/M
```

```
## region
##     1     2     3     4     5
## 0.159 0.201 0.245 0.235 0.160
```

Signal Detection Model (Equal Variancess)



Your Turn

- Using `R`, code up a simulation model for both new and old items.
- Use your simulation code to predict the probability of any response for any set of four bounds for both types of stimuli?

Calculating Probabilities

- `pnorm()` is area to the left
- sure new: `pnorm(bound[1],0,1)`
- maybe new: `pnorm(bound[2],0,1)-pnorm(bound[1],0,1)`

Your Turn

Use `pnorm()` rather than simulation to compute all probabilities for both item types.

My Turn

```r
mu=1
bounds=seq(-1,1,length=4)
cumNew=c(0,pnorm(bounds,0,1),1)
pNew=diff(cumNew)
cumOld=c(0,pnorm(bounds,mu,1),1)
pOld=diff(cumOld)
```

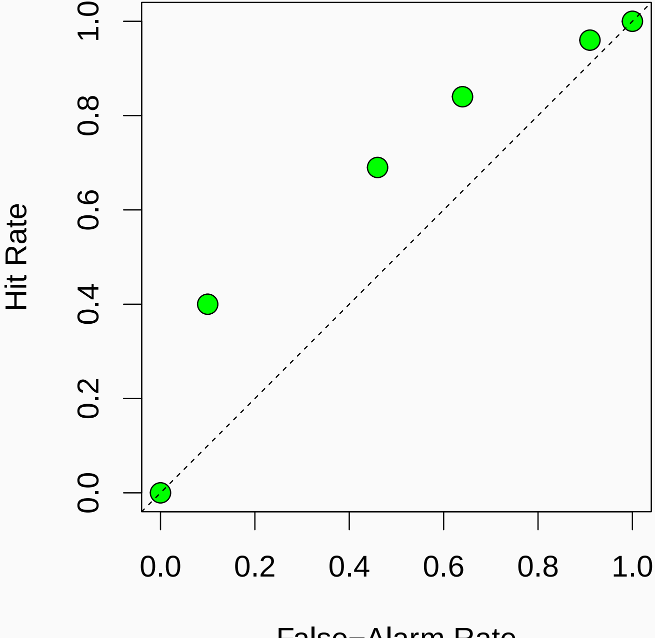
Data Visualization

	Sure New	Maybe New	Unsure	Maybe Old	Sure Old
Old Items	4	12	15	29	40
New Items	9	27	18	36	10

	Sure New	Maybe New	Unsure	Maybe Old	Sure Old
Old Items	0.04	0.12	0.15	0.29	0.4
New Items	0.09	0.27	0.18	0.36	0.1

	Sure New	Maybe New	Unsure	Maybe Old	Sure Old
Old Items	0.96	0.84	0.69	0.4	0
New Items	0.91	0.64	0.46	0.1	0

Data Visualiztion, ROC



Your Turn

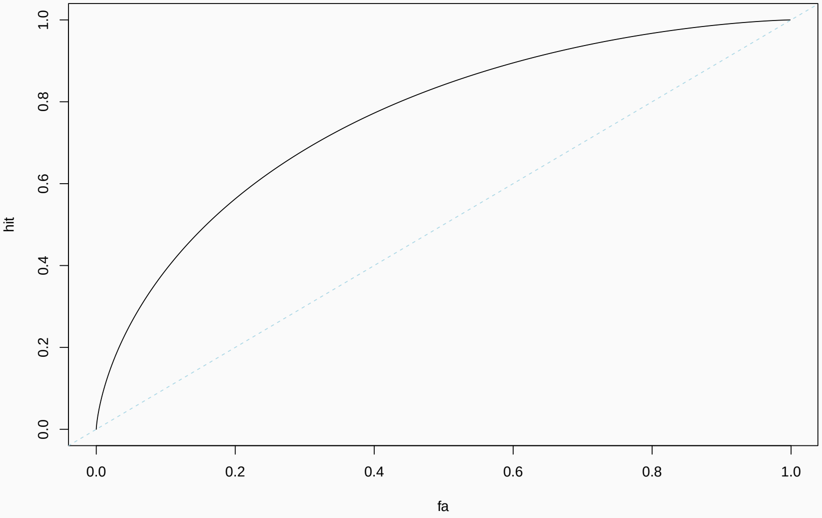
Use R and your previous code to explore what types of ROC curves the signal detection model predicts.

- HINT, use a large number of bounds rather than just 4. This allows you to draw the ROC curves in more detail.

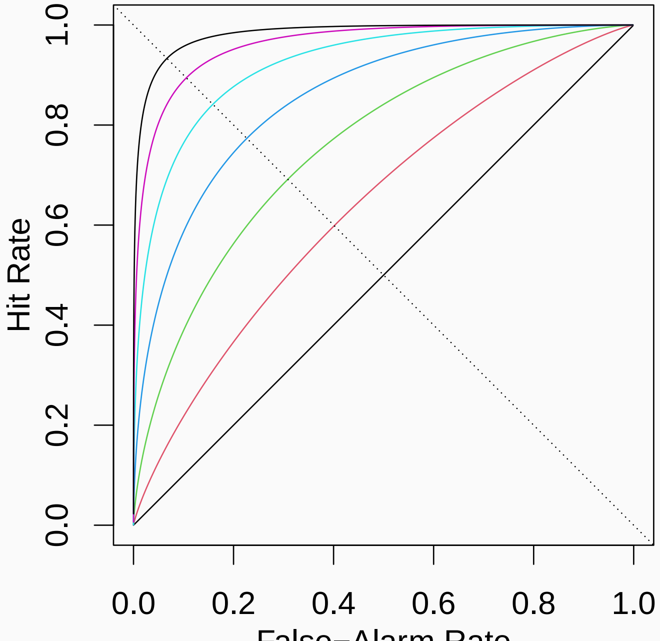
My Turn

```
crit=seq(-3,5,.01)
hit=1-pnorm(crit,1)
fa=1-pnorm(crit)
plot(fa,hit,typ='l')
abline(0,1,lty=2,col='lightblue')
```

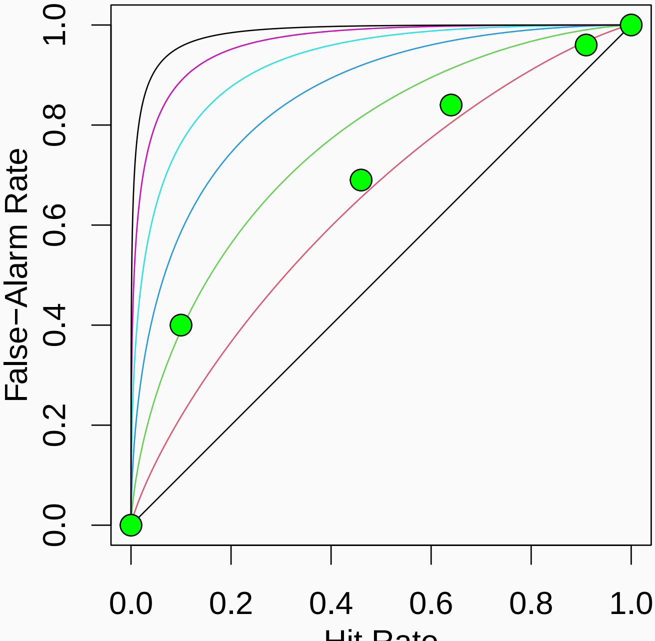
My Turn



My Turn



Curves + Data



Your Turn

What is a good notation to formally express the signal detection model? (this is hard)

- Parameters:
  - center for old items, μ
  - criteria $\gamma_1 \leq \gamma_2 \leq \gamma_3 \leq \gamma_4$
- Data
  - N_{ij} is number of responses for the ith item, jth response category, $i = 1, 2$, $j = 1, \ldots, 5$.
  - $\boldsymbol{N}_i = (N_{i1}, \ldots, N_{i5})$

My Turn

Let $S_i = \sum_j N_{ij}$ be the number of trials in the old and new item conditions for $i = 1, 2$, respectively. Then, we may model the vector of cell counts per item type as

$$\boldsymbol{N}_i \sim \text{Multinomial}(\boldsymbol{p}_i, S_i),$$

where $\boldsymbol{p}_i = (p_{i1}, \ldots, p_{i5})$ is a vector of probabilities for each item condition subject to the constraint $\sum_j p_{ij} = 1$.

My Turn

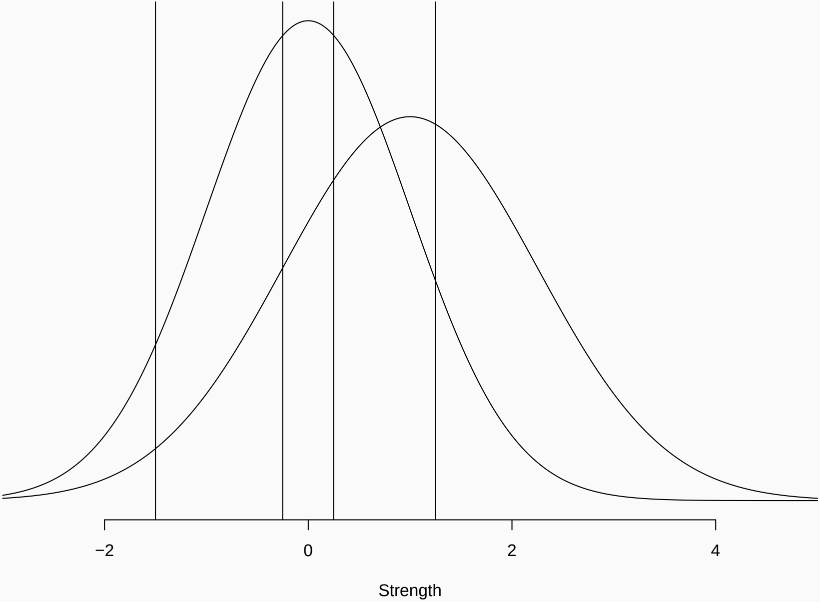
- $\Phi(x)$ is area to the left of x for a standard normal
- $\Phi\left(\frac{x-m}{s}\right)$ is area to the left of x for a normal with mean m and variance s^2.

My Turn

$$
\begin{aligned}
p_{1j} &= \Phi(\gamma_j) - \Phi(\gamma_{j-1}), \\
p_{2j} &= \Phi\left(\gamma_j - \mu\right) - \Phi\left(\gamma_{j-1} - \mu\right),
\end{aligned}
$$

where $\gamma_0 = -\infty$ and $\gamma_5 = \infty$.

Unequal Variance Signal Detection Model



Your Turn

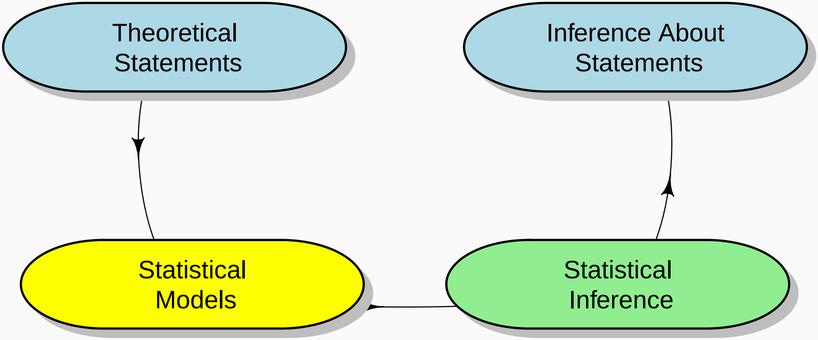
- Write code to compute hit and false alarm rates for a set of parameters.
- Draw some ROC curves
- Write the formal model

Is This a Good Model?

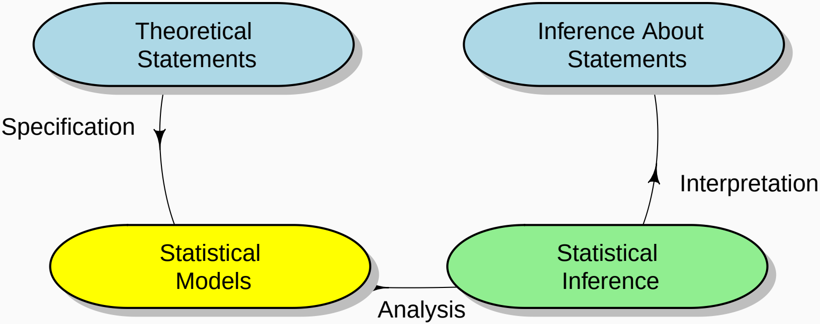
Good?:

- true?
- accurate?
- useful?
- insightful?
- falsifiable?

Overview of Modeling



Overview of Modeling



Theoretical Statements

Examples:

- Working-Memory Training Transfer: *"Practicing one working-memory task will increase performance in other related tasks."*
- Weber's Law: $\Delta I = kI$.
- Stroop: *Performace to congruent items is better than to incongruent ones*
- A certain drug is beneficial to all but benefits women more than men

Theoretical Statments

Theoretical Statments

- Each statement describes constraint in the world

Theoretical Statments

- Each statement describes constraint in the world
- None of these statements may be tested with data

Theoretical Statments

- Each statement describes constraint in the world
- None of these statements may be tested with data
- None has specification of sampling noise or errors in measurements.

Theoretical Statments

- Each statement describes constraint in the world
- None of these statements may be tested with data
- None has specification of sampling noise or errors in measurements.
- And that is appropriate

Models

Models Are Instantiations of Theories That Predict Data

Models

Models Are Instantiations of Theories That Predict Data

1. Models are proxies that capture the constraint in a theory.

Models

Models Are Instantiations of Theories That Predict Data

1. Models are proxies that capture the constraint in a theory.
2. Models lead to probability statements on observables

Are Models Really False or True?

George Box famously said, "*All models are false, but some models are more useful than others.*"

Are Models Really False or True?

George Box famously said, "*All models are false, but some models are more useful than others.*"

- Box was trying to get us to take a stand against truth

Are Models Really False or True?

George Box famously said, "*All models are false, but some models are more useful than others.*"

- Box was trying to get us to take a stand against truth
- I would rather take a stand against truth and falseness

Are Models Really False or True?

George Box famously said, "*All models are false, but some models are more useful than others.*"

- Box was trying to get us to take a stand against truth
- I would rather take a stand against truth and falseness
- Box should have said, *All models are not even false. Some models are more useful than others.*"

Is This Metro Map True or False?

S

Is This Metro Map True or False?

S

- Map captures the order of stations and intersections of lines

Is This Metro Map True or False?

s

- Map captures the order of stations and intersections of lines
- Map distorts distances and colors.

Models Are Neither True Nor False

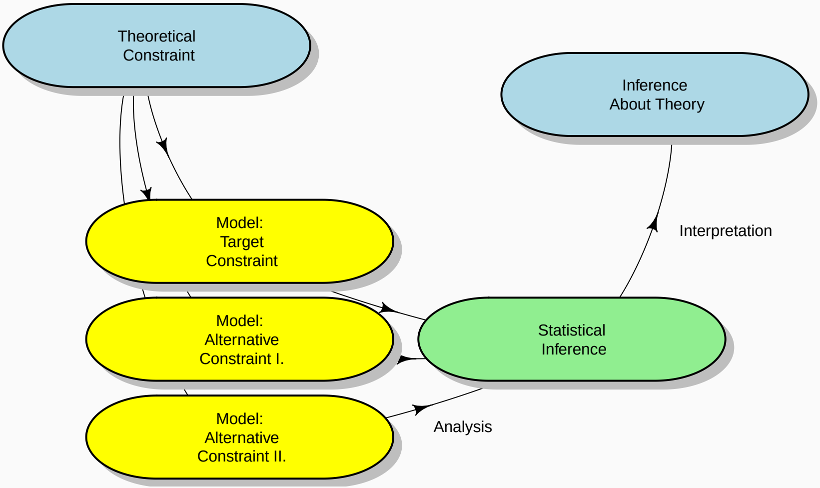
Models Capture Theoretically Meaningful Constraint in Data

Models Are Neither True Nor False

An Agenda of Comparison

- Instantiate a theory as a model that places constraint on data
- Instantiate a theoretically meaningful alternatives as an alternative models. These alternative models should place contrasting constraint on data.
- Compare the *relative evidence* for the target model against the alternatives.

Overview of Modeling (Revised)



Your Turn

- What are the theoretical statements / constraints in the theory of signal detection?
- What is the role of the normal? Is it like color on tracks or like order of stations? That is, does it have psychological content?