Multilevel/Hierarchical Models II

Lim Wai Yee, *The Hanging Gardens of Babylon*

SML480: Pedagogy of Data Science, Spring 2020

Michael Guerzhoy

Item-response (Rasch) model

- J persons, K items
- $y_i = 1$ if the response is correct
- Model: $P(y_i = 1) = \sigma(\alpha_{j[i]} - \beta_{k[i]})$
  - a_j is the ability of person j
  - β_k is the difficulty of problem k
- Non-identifyable: can increase the alphas and the betas by a constant and get the same probabilities
  - Can subtract the mean alpha to deal with this

Multilevel model

- $\alpha_j \sim N(0\ , \sigma_\alpha^2)$
- $\beta_j \sim N(\mu_\beta, \sigma_\beta^2)$
- μ_α set to 0 to avoid non-identifyability

Item specific "discrimination" parameter

- $P(y_i = 1) = \sigma(\gamma_{k[i]}(\alpha_{j[i]} - \beta_{k[i]}))$

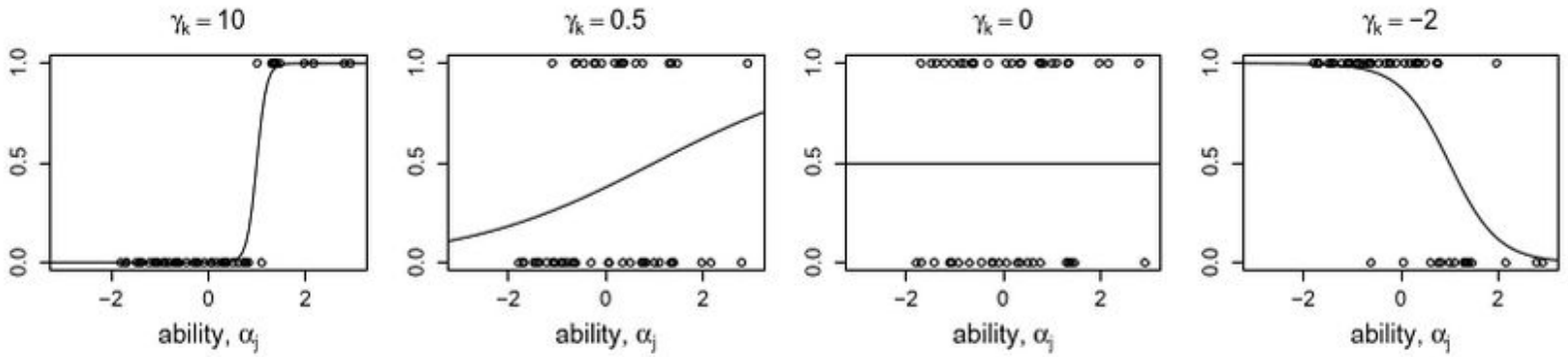


Figure 14.14 *Curves and simulated data from the logistic item-response (Rasch) model for items k with "difficulty" parameter* $\beta_k = 1$ *and high, low, zero, and negative "discrimination" parameters* γ_k.

- $P(y_i = 1) = \sigma(\gamma_{k[i]}(\alpha_{j[i]} - \beta_{k[i]}))$
- Identifyability problems?

Stroop task


Stroop Task 1
READ THE WORDS
RED BLUE
BLUE RED
GREEN GREEN
RED GREEN
GREEN BLUE
BLUE GREEN
Stroop Task 2
SAY THE COLOUR OF THE INK
XXXX XXXX
XXXX XXXX
XXXX XXXX
XXXX XXXX
XXXX XXXX
XXXX XXXX
Stroop Task 3
SAY THE COLOUR OF THE INK
RED BLUE
BLUE RED
GREEN GREEN
RED GREEN
GREEN BLUE
BLUE GREEN

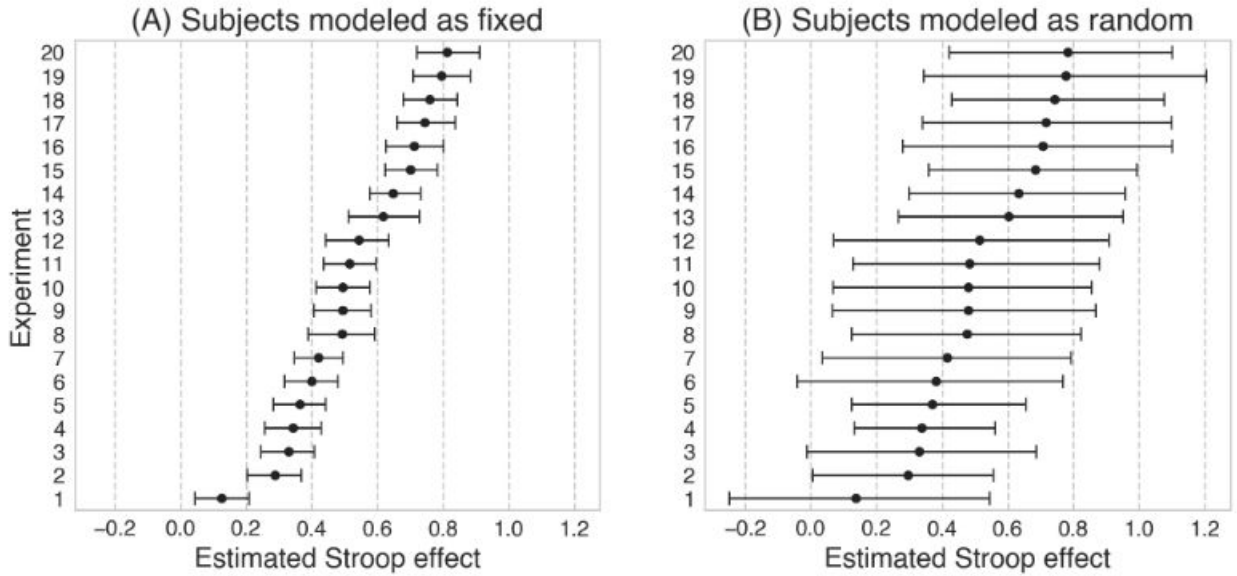
Strawman model

- $y_{ij} = \beta_0 + \beta_1 X_{ij} + e_{ij}, e_{ij} \sim N(0, \sigma_e^2)$
  - y_{ij}: reaction time of of i-th subject, j-trial
  - X_{ij}: congruent/incongruent condition for i-th subject j-th trial

Model 2

- $y_{ij} = \beta_0 + \beta_1 X_{ij} + u_{0i} + u_{1i} X_{ij} + e_{ij}$

$$u_{0i} \sim N(0, \sigma^2_{u0})$$
$$u_{1i} \sim N(0, \sigma^2_{u1})$$
$$e_{ij} \sim N(0, \sigma^2_e)$$



Yarkoni's argument #1

- In the first model, rejecting $\beta_1 = 0$ means that it is unlikely *for the particular subjects we observe* that there was no difference between congruent/non-congruent conditions
- In the second model, rejecting $\beta_1 = 0$ means that for subjects as modeled by Model 2, it is unlikely that there was no difference between congruent/non-congruent conditions

Yarkoni's argument #2

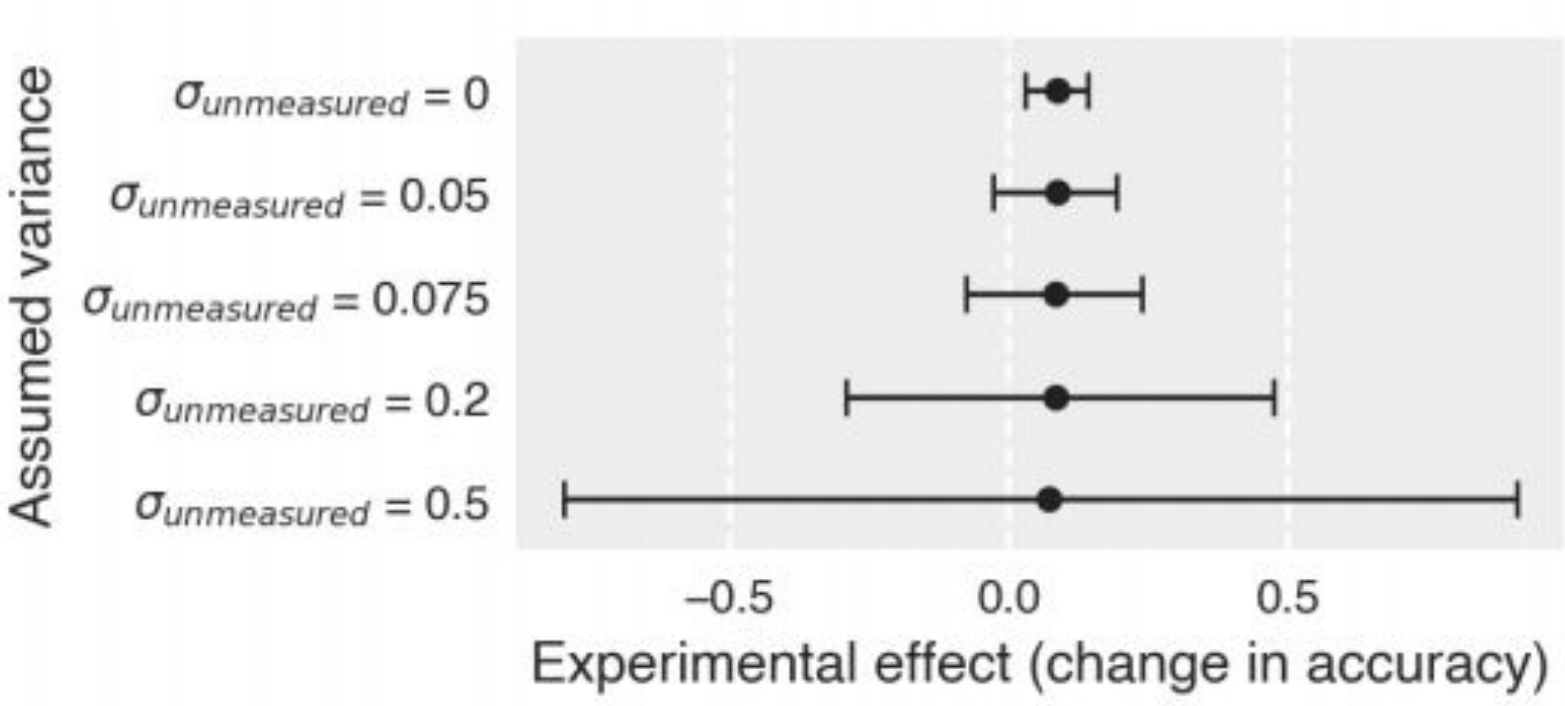
- Research subjects are not the only random effects: so are stimuli, experimenters, research sites, etc.
- Stimuli as non-random effects
  - Strictly speaking, any specific experiment shows that *the particular stimuli used* have an effect

The effect of stimuli as random

?

$$
\begin{aligned}
y_{ps} &= \beta_0 + \beta_1 X_{ps} + u_{0s} + u_{1s} X_{ps} + u_2 X_{ps} + e_{ps} \\
u_{0s} &\sim \mathcal{N}(0, \sigma^2_{u_0}) \\
u_{1s} &\sim \mathcal{N}(0, \sigma^2_{u_1}) \\
u_2 &\sim \mathcal{N}(0, \sigma^2_{u_2}) \\
e_{ps} &\sim \mathcal{N}(0, \sigma^2_e)
\end{aligned}
\tag{4}
$$

Here, p indexes participants, s indexes sites, X_{ps} indexes the experimental condition assigned to participant p at site s, the β terms encode the fixed intercept and condition slope, and the u terms encode the random effects (site-specific intercepts u_0, site-specific slopes u_1, and the stimulus effect u_2). The novel feature of this model is the inclusion of u_2, which would ordinarily reflect the variance in outcome associated with random stimulus sampling, but is constant in our dataset (because there's only a single stimulus).


Assumed variance
$\sigma_{unmeasured} = 0$
$\sigma_{unmeasured} = 0.05$
$\sigma_{unmeasured} = 0.075$
$\sigma_{unmeasured} = 0.2$
$\sigma_{unmeasured} = 0.5$
–0.5
0.0
0.5
Experimental effect (change in accuracy)

Are subjects modeled correctly?

“Existence proof”

- Any particular study with a significant effect can be treated as evidence that a an interesting effect is observed *under some circumstances*
  - The Stanford prison experiment https://www.vox.com/2018/6/13/17449118/stanford-prison-experiment-fraud-psychology-replication
  - Milgram’s electroshock test
- Less famous and more quantitative examples?
- https://www.pnas.org/content/111/23/8410 ?

Where to go from here?

- “Do something else”
- “Embrace qualitative analysis”
- “Adopt better standards”
  - So, is this all about rhetoric in the abstract?
- Fit more expansive models + design with variation in mind
- Make riskier predictions
  - Stop affirming the consequent
-