So simple, yet so effective

Ottavia M. Epifania
University of Trento
ottavia.epifania@unitn.it

Beyond Summer School

May, 30, 2025

How it started:

A random lesson by Professor Pastore on Generalized Linear Mixed Effects Models

How it started:

A random lesson by Professor Pastore on Generalized Linear Mixed Effects Models

How it ended:

American Psychological Association
Psychological Methods

ISSN: 1082-989X
https://doi.org/10.1037/met0000708

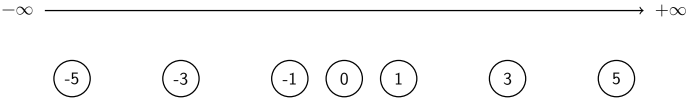
A Guided Tutorial on Linear Mixed-Effects Models for the Analysis of Accuracies and Response Times in Experiments With Fully Crossed Design

Ottavia M. Epifania, Pasquale Anselmi, and Egidio Robusto
Department of Philosophy, Sociology, Education and Applied Psychology, University of Padova

https://doi.org/10.1037/met0000708

Fully-crossed structures

An example: The SNARC effect



Small numbers:
Perceived on the left

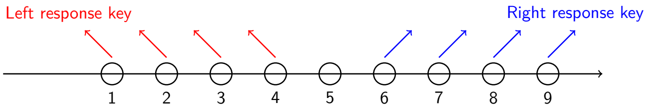
Large numbers:
Perceived on the right

A *sample* of small numbers:
$1, 2, 3, 4$

A *sample* of large numbers:
$6, 7, 8, 9$

Two conditions:

The “natural” one (so-called *compatible* condition)

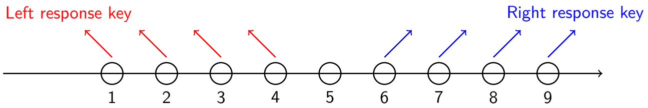


A *sample* of small numbers:
$1, 2, 3, 4$

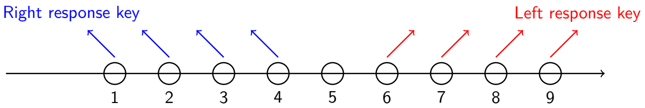
A *sample* of large numbers:
$6, 7, 8, 9$

Two conditions:

The "natural" one (so-called *compatible* condition)



The "innatural" one (so-called *incompatible* condition)



$t = \{1, 2, \dots, T\}$: Number of trials (condition $\times$ stimulus $\times$ respondent)

		Small Numbers				Large Numbers			
	Condition	1	2	3	4	6	7	8	9
Jane	Compatible	y_{cj1}	y_{cj2}	y_{cj3}	y_{cj4}	y_{cj6}	y_{cj7}	y_{cj8}	$\sum_{t=1}^{T} y_{cj}/T$
	Incompatible	y_{ij1}	y_{ij2}	y_{ij3}	y_{ij4}	y_{ij6}	y_{ij7}	y_{ij8}	$\sum_{t=1}^{T} y_{ij}/T$
Mario	Compatible	y_{cm1}	y_{cm2}	y_{cm3}	y_{cm4}	y_{cm6}	y_{cm7}	y_{cm8}	$\sum_{t=1}^{T} y_{cm}/T$
	Incompatible	y_{im1}	y_{im2}	y_{im3}	y_{im4}	y_{im6}	y_{im7}	y_{im8}	$\sum_{t=1}^{T} y_{im}/T$

Person-level scores

$$s_p = \frac{\bar{X}_{p,\text{comp}} - \bar{X}_{p,\text{inc}}}{sd_{\text{pooled}}}$$

Person-level scores

$$s_p = \frac{\bar{X}_{p,\text{comp}} - \bar{X}_{p,\text{inc}}}{sd_{\text{pooled}}}$$

Advantages

Ease of computation
Ease of interpretation

Person-level scores

$$s_p = \frac{\bar{X}_{p,\text{comp}} - \bar{X}_{p,\text{inc}}}{sd_{\text{pooled}}}$$

Advantages

Ease of computation
Ease of interpretation

(Implicit) Assumptions

1. Being slow (less accurate) in one condition = being fast (or more accurate) in the opposite one: 0 means absence of bias
2. All stimuli have the same impact (fixed effects)

A long tradition

Respondents are random factors

Sampled from a larger population
Need for acknowledging the sampling variability
Results can be generalized to other respondents belonging to the same population

A long tradition

Respondents are random factors

Sampled from a larger population
Need for acknowledging the sampling variability
Results can be generalized to other respondents belonging to the same population

Stimuli/items are fixed factors

Taken to be entire population
There is no sampling variability
There is no need to generalize the results because the stimuli are the population

With long lasting consequences

- Generalization of the results is impaired
- Error variance everywhere, left free to bias everything
- The information at the stimulus level is lost

With long lasting consequences

- Generalization of the results is impaired
- Error variance everywhere, left free to bias everything
- The information at the stimulus level is lost

Σ	ψ
Linear Mixed Effects Models	Rasch model

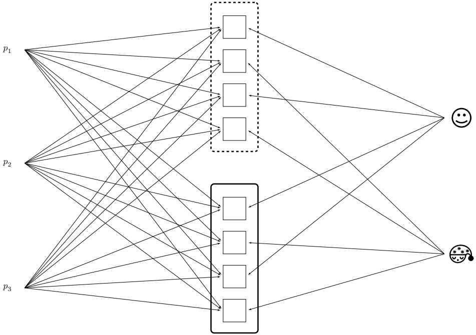
With long lasting consequences

- Generalization of the results is impaired
- Error variance everywhere, left free to bias everything
- The information at the stimulus level is lost

$\sum$	ψ
Linear Mixed Effects Models	Rasch model

Rasch-like parametrization estimated with Linear Mixed Effects Models

When?



Sample-level differences:
Compatible and incompatible can be defined *a priori*
(SNARC effect)

Individual differences:
Compatible and incompatible are defined within each respondent
(Implicit Association Test)

A Classic of Psychometrics

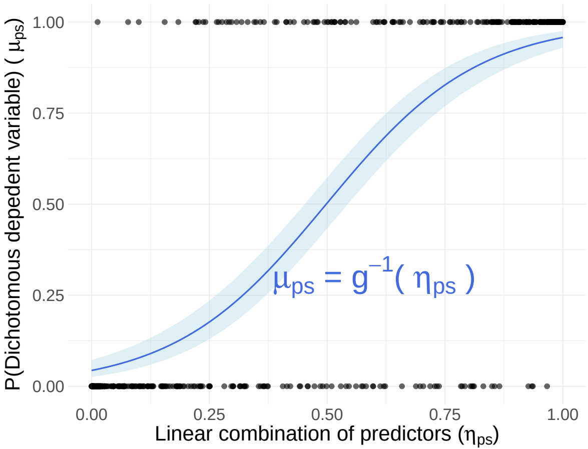
The Rasch Model

$$P(x_{ps} = 1|\theta_p, b_s) = \frac{\exp(\theta_p - b_z)}{1 + \exp(\theta_p - b_z)}$$

θ_p: Latent trait of person p

b_s: “challenging” power of stimulus s

A GLM for dichotmous responses



Logit link function g:

$$g(\eta_{ps}) = log\left(\frac{\mu_{ps}}{1-\mu_{ps}}\right)$$

Inverse g^{-1}

$$g^{-1} = \frac{\exp(\eta_{ps})}{1+\exp(\eta_{ps})}$$

The log-normal model

$$E(t_{ps}|\tau_p, \delta_s) = \delta_s - \tau_p + \varepsilon$$

τ_p: the speed of person p

δ_s: the time intensity of stimulus s

The log-normal model

$$E(t_{ps}|\tau_p, \delta_s) = \delta_s - \tau_p + \varepsilon$$

τ_p: the speed of person p

δ_s: the time intensity of stimulus s

A linear model with an identity function!

In summary

Rasch

$$P(x_{ps} = 1) = \frac{\exp(\theta_p - b_z)}{1 + \exp(\theta_p - b_z)}$$

GLM (inverse function)

$$P(x_{ps} = 1) = \frac{\exp(\theta_p + b_s)}{1 + \exp(\theta_p + b_s)}$$

Log-normal

$$E(t_{ps}|\tau_p, \delta_s) = \delta_s - \tau_p + \varepsilon$$

LM (identity function)

$$E(t_{ps}|\tau_p, \delta_s) = \delta_s + \tau_p + \varepsilon$$

Random Factors and Effects

In a LM:

$$\eta = \mathbf{X}\beta$$

$\mathbf{X}$: Model Matrix

β: Coefficients

In a LM:

$$\eta = \mathbf{X}\beta$$

$\mathbf{X}$: Model Matrix

β: Coefficients

Needs to be extended:

$$\eta = \mathbf{X}\beta + Zd$$

d: Random effects associated to the random factors in Z ... Not model parameters! *Best Linear Unbiased Predictors*

Γ: Parameters estimated for the random factors in the model (variances and

The maximal model

Address all the possible sources of random variability that can be expected

The maximal model

Address all the possible sources of random variability that can be expected

The models that are useful for ones aim

The maximal model

Address all the possible sources of random variability that can be expected

The models that are useful for ones aim

Common goal: Investigate the changes in the performance of the respondents between the associative conditions

The maximal model

Address all the possible sources of random variability that can be expected

The models that are useful for ones aim

Common goal: Investigate the changes in the performance of the respondents between the associative conditions

Less common: Investigate the changes in the functioning of the stimuli between the associative conditions

Preliminarities

Index	Meaning	Variable
$p = 1, \dots, P$	Respondent	`respondents`
$s = 1, \dots, S$	Stimulus	`stimuli`
$c \in \{0, 1\}$	Associative condition	`condition`
i	Trial	

Accuracy:
GLMM
$y = [0, 1]$

Log-time response
LMM
$y = [0, +\infty)$ (log-transformed)

$$\varepsilon \sim \mathcal{N}(0, \sigma^2)$$

Model 1

Mathematical Notation

$$y = \beta_c X_c + \alpha_p[i] + \alpha_s[i]$$

lme4 notation

```
y ~ 0 + condition + (1|stimuli) + (1|respondents)
```

Rasch-like parametrization

	GLMM	LMM
respondents	θ_p	τ_p
stimuli	b_s	δ_s

Model 2

Mathematical Notation

$$y = \beta_c X_c + \alpha_p[i] + \beta_s[i] c_i$$

lme4 notation

```
y ~ 0 + condition + (0+condition|stimuli) + (1|respondents)
```

Rasch-like parametrization

	GLMM	LMM
respondents	θ_p	τ_p
stimuli	b_{sc}	δ_{sc}

Model 3

Mathematical Notation

$$y = \beta_c X_c + \beta_p[i] c_i + \alpha_s[i]$$

lme4 notation

```
y ~ 0 + condition + (1|stimuli) + (0+condition|respondents)
```

Rasch-like parametrization

	GLMM	LMM
respondents	θ_{pc}	τ_{pc}
stimuli	b_s	δ_s

All models are wrong...

Find the useful model via model comparison: AIC and BIC

The lower the value, the better the model

! AIC, BIC, and model complexity:

Total number of parameters: β and Γ
NOT the levels in d

Model 2 and Model 3: Same complexity, different focus

The chosen model is the least wrong model *given the models considered*:
Relativity applies everywhere

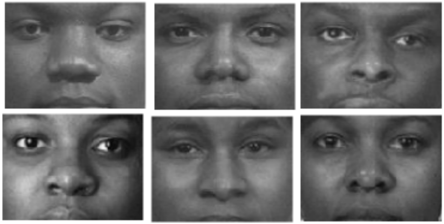
Real data

The Implicit Association Test

12 Object stimuli

White people faces

Black people faces



16 Attribute stimuli

Positive attributes

Good, laughter, pleasure, glory, peace, happy, joy, love

Negative attributes

Evil, bad, horrible, terrible, nasty, pain, failure, hate

Get set

```
# install package for fitting lmms
install.packages("lme4")
# nice plots :)
install.packages("ggplot2")


library(lme4)
library(ggplot2)
```

The data

```r
data = read.csv("data/example-data.csv",
                header = TRUE, sep =",")
head(data)
```

```
  respondent condition  stimuli accuracy latency
1          1 Whitegood     hate        1    1224
2          1 Whitegood     bf14        1    5160
3          1 Whitegood laughter        1    1214
4          1 Whitegood     bf56        1    1143
5          1 Whitegood     evil        1     827
6          1 Whitegood      wf3        1    1859
```

Number of trials × condition × respondent:

```
table(data$respondent, data$condition)
```

```
     Whitebad Whitegood
  1        60        60
  2        60        60
  3        60        60
  4        60        60
  5        60        60
  6        60        60
....
```

Model 1: $y_i = logit^{-1}(\alpha + \beta_c X_c + \alpha_{p[i]} + \alpha_{s[i]})$

```
accuracy1 = glmer(accuracy ~ 0 + condition + (1|stimuli) + (1|respondent),
                  data = data,
                  family = "binomial")
```

Model 2: $y_i = logit^{-1}(\alpha + \beta_c X_c + \alpha_{p[i]} + \beta_{s[i]} c_i)$

```
accuracy2 = glmer(accuracy ~ 0 + condition + (0 + condition|stimuli) +
                    (1|respondent),
                  data = data,
                  family = "binomial")
```

Model 3: $y_i = logit^{-1}(\alpha + \beta_c X_c + \alpha_{s[i]} + \beta_{p[i]} c_i)$

```
accuracy3 = glmer(accuracy ~ 0 + condition + (1|stimuli) +
                    (0 + condition|respondent),
                  data = data,
                  family = "binomial")
```

LMMs for log-time responses

Model 1: $y_i = \alpha + \beta_c X_c + \alpha_{p[i]} + \alpha_{s[i]} + \varepsilon_i$

```
logtime1 = lmer(log(latency) ~ 0 + condition + (1|stimuli) + (1|respondent),
                data = data,
                REML = FALSE)
```

Model 2: $y_i = \alpha + \beta_c X_c + \alpha_{p[i]} + \beta_{s[i]} c_i + \varepsilon_i$

```
logtime2 = lmer(log(latency) ~ 0 + condition + (0 + condition|stimuli) +
                  (1|respondent),
                data = data,
                REML = FALSE)
```

Model 3: $y_i = \alpha + \beta_c X_c + \alpha_{s[i]} + \beta_{p[i]} c_i + \varepsilon_i$

```
logtime3 = lmer(log(latency) ~ 0 + condition + (1|stimuli) +
                  (0 + condition|respondent),
                data = data,
                REML = FALSE)
```

GLMMs

⚠ Important!

The use of the anova() function is just for the convenience of having all the information on the same page!

```
anova(accuracy1, accuracy2, accuracy3)
```

```
Data: data
Models:
accuracy1: accuracy ~ 0 + condition + (1 | stimuli) + (1 | respondent)
accuracy2: accuracy ~ 0 + condition + (0 + condition | stimuli) + (1 | respondent)
accuracy3: accuracy ~ 0 + condition + (1 | stimuli) + (0 + condition | respondent)
          npar    AIC    BIC  logLik -2*log(L)  Chisq Df Pr(>Chisq)
accuracy1    4 4144.3 4172.1 -2068.1    4136.3
accuracy2    6 4141.3 4183.1 -2064.7    4129.3 6.9271  2    0.03132 *
accuracy3    6 4145.2 4187.0 -2066.6    4133.2 0.0000  0
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

LMMs

Important!

The use of the `anova()` function is just for the convenience of having all the information on the same page!

```
anova(logtime1, logtime2, logtime3)
```

```
Data: data
Models:
logtime1: log(latency) ~ 0 + condition + (1 | stimuli) + (1 | respondent)
logtime2: log(latency) ~ 0 + condition + (0 + condition | stimuli) + (1 | respondent)
logtime3: log(latency) ~ 0 + condition + (1 | stimuli) + (0 + condition | respondent)
         npar    AIC    BIC  logLik -2*log(L)   Chisq Df Pr(>Chisq)
logtime1    5 6073.2 6108.0 -3031.6    6063.2
logtime2    7 6061.4 6110.2 -3023.7    6047.4  15.743  2  0.0003815 ***
logtime3    7 5657.2 5705.9 -2821.6    5643.2 404.249  0
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Useful Models:

GLMMs
Model 2
θ_p
b_{WGBB} and b_{BGWB}

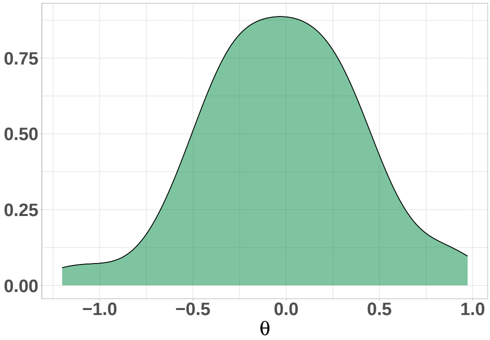
The IAT effect is mostly due to variations in the *stimuli functioning* between conditions, while the performance of the respondents seems unaltered

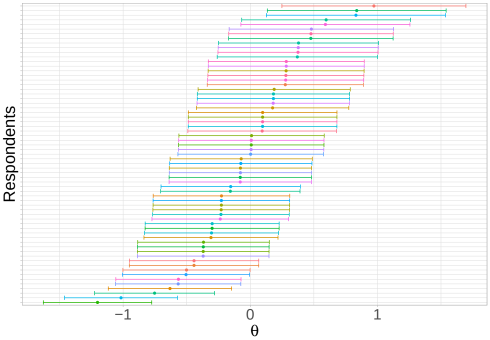
LMMs
Model 3
τ_{WGBB} and τ_{BGWB}
δ_s

The IAT effect is mostly due to variations in the *performance of the respondents* between conditions, while the functioning of the stimuli appears not affected

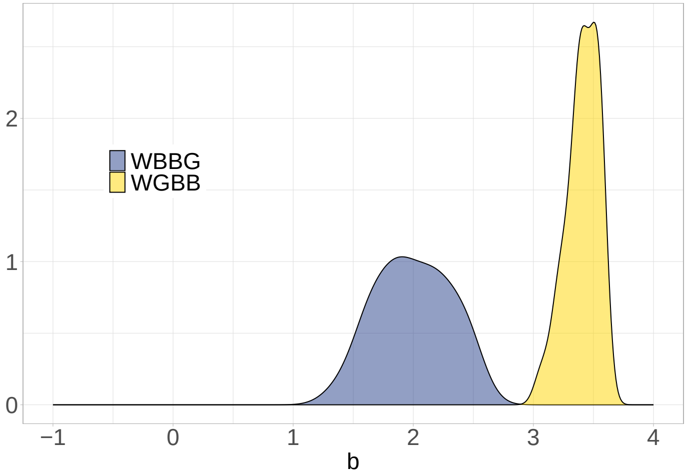
Results should be interpreted together!

θ_p

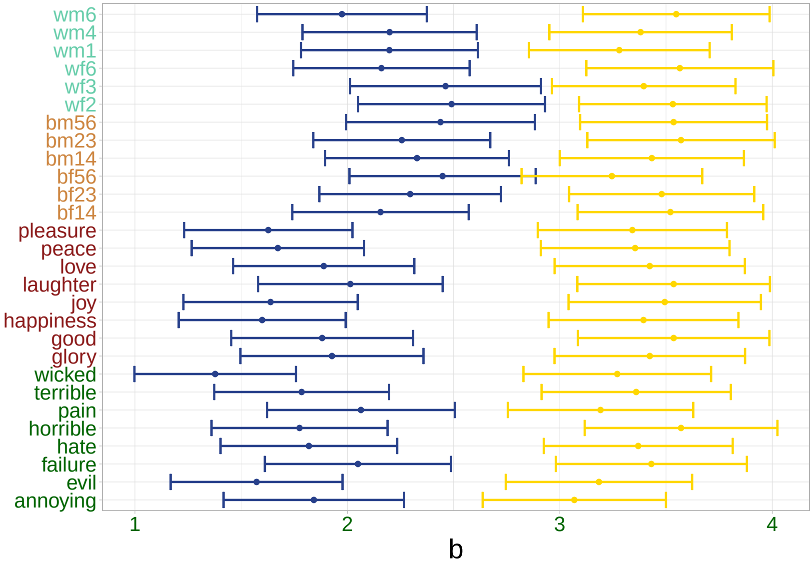




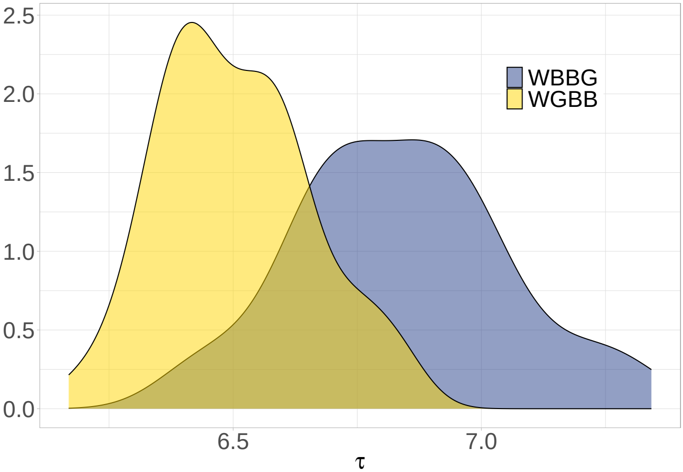
b_{WGBB} and b_{WGBB}



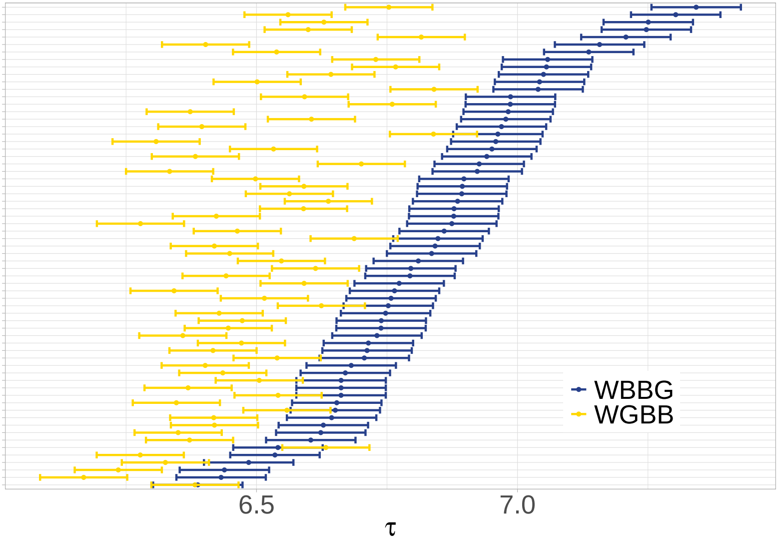
Rasch-like estimates



$\tau_{\mathbf{WGBB}}$ and $\tau_{\mathbf{BGWB}}$

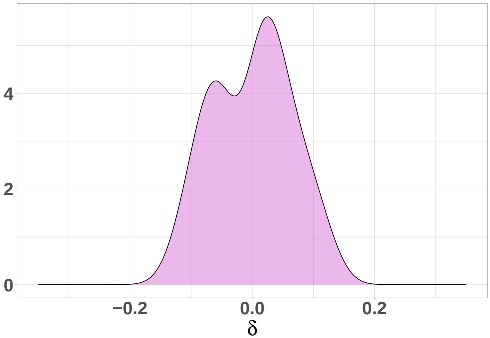


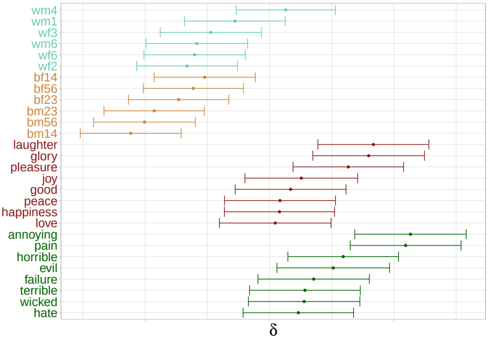
Log-normal estimates



Log-normal estimates

δ_s





The end

- The best model depends on the other models... sometimes useful, never right
- The sky is the limit... but do not over complicate things

- The best model depends on the other models... sometimes useful, never right
- The sky is the limit... but do not over complicate things

HOWEVER

- Time and accuracy are independent from one another, pretty bold assumption