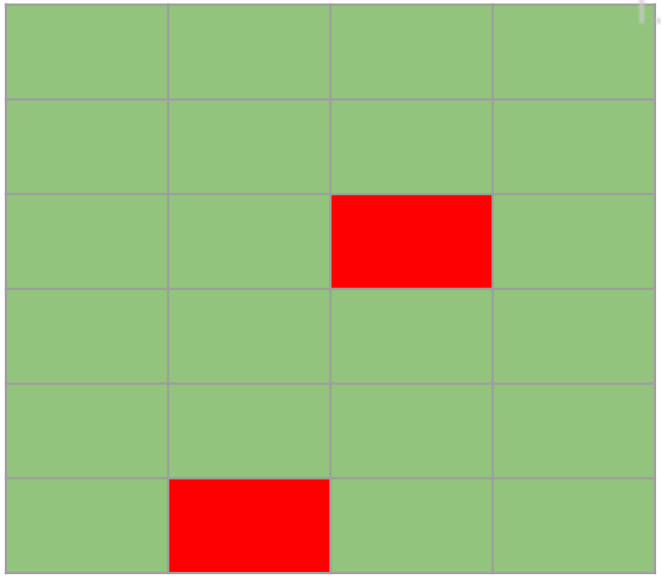
# The Hitchhiker's Guide to Mixed Models

Teresa Del Bianco

02 October 2020

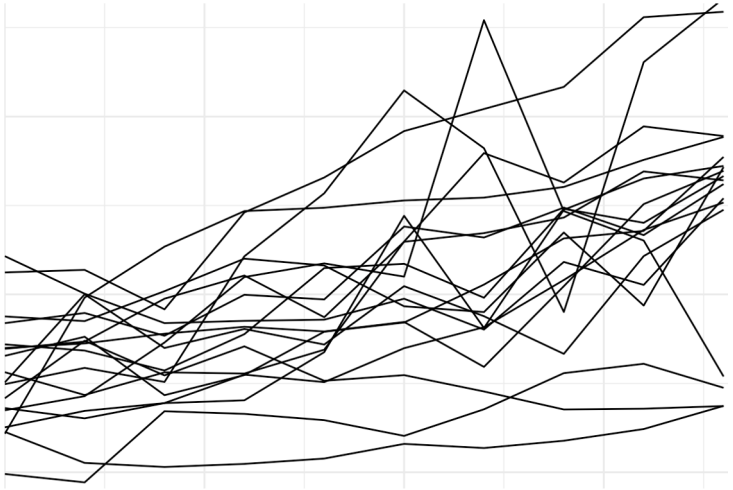
# How I got to mixed models

Final year of PhD





Job as biostatistician



# What's in a name?

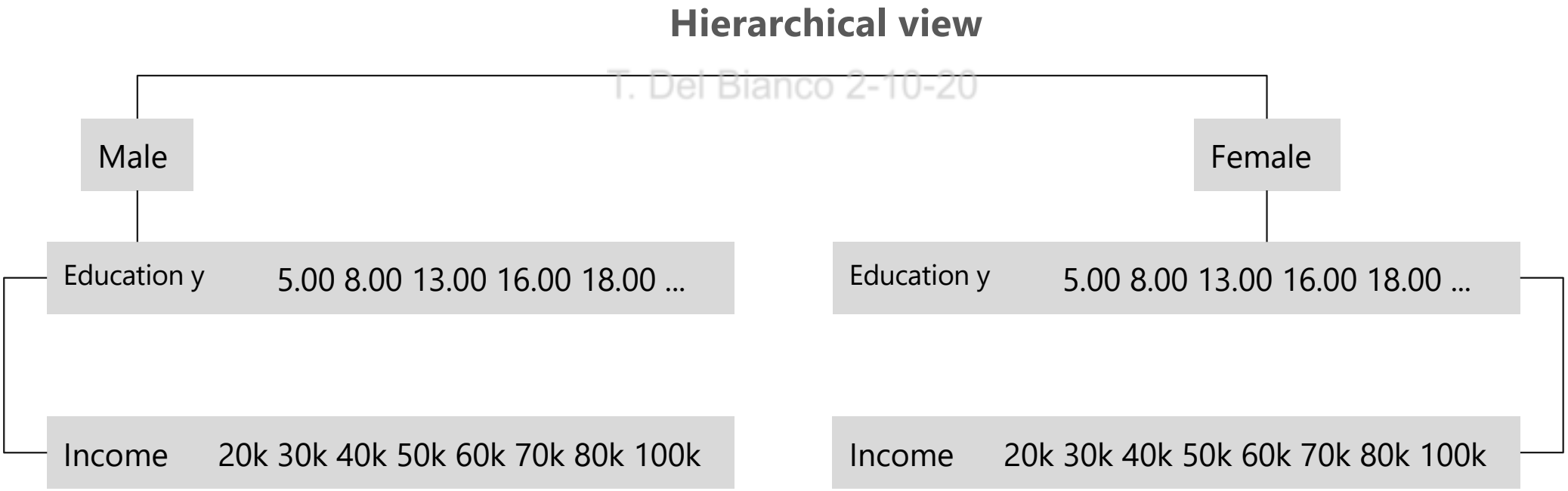
Multilevel Model = parameters vary at more than one level



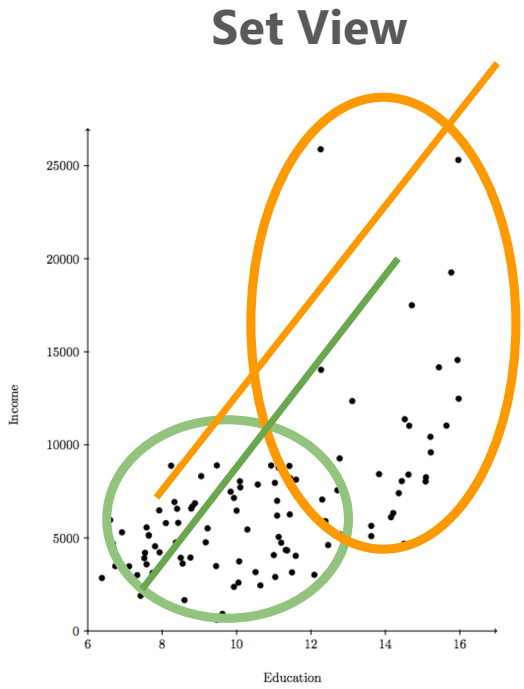
hierarchical linear models
linear mixed-effect model
mixed models
nested data models
random coefficient
random-effects models
random parameter models
split-plot design

→ **MIXED**

# Varying on more than 1 level: nesting



# Varying on more than 1 level: varying coefficients





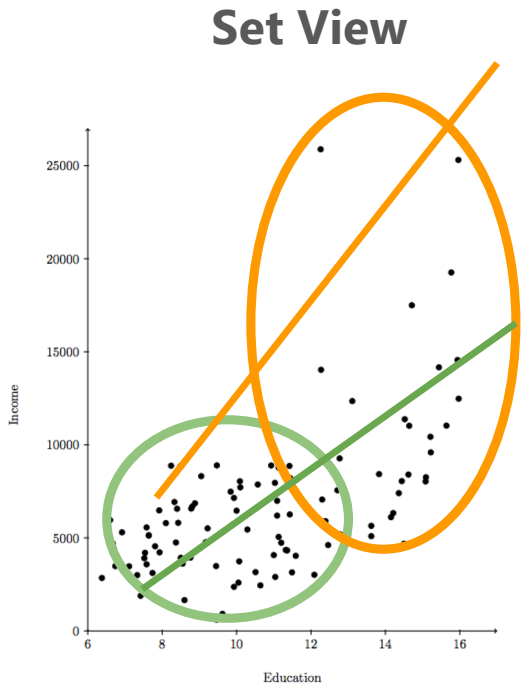

**Formula:**

$$y_i = B_0 + B_1X_i + B_{0i} + E_i$$

Varying **Intercept** per Level

Fixed Effects | Random Effect

# Varying on more than 1 level: varying coefficients






**Formula:**

$$y_i = B_0 + B_1X_i + B_{0i} + B_{1i}X_i + E_i$$

Varying **Intercept and Slope** per Level

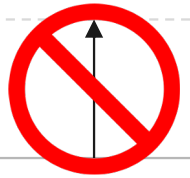
Random Effect

# Much Ado About Nothing?

hierarchical linear models
linear mixed-effect model
mixed models
nested data models
random coefficient
random-effects models
random parameter models
split-plot design



"Mixed" (repeated measures) Anova
Generalised Estimating Equations (GEE)



Two types of Effects:

Fixed
Random

→ **MIXED** →

Used for nested designs (e.g., 2+ groups, within subject exp conditions, longitudinal, violates assumptions) where the observations are **repeated** and, in most cases, **unbalanced**

# Different ways of dealing with repeated measures

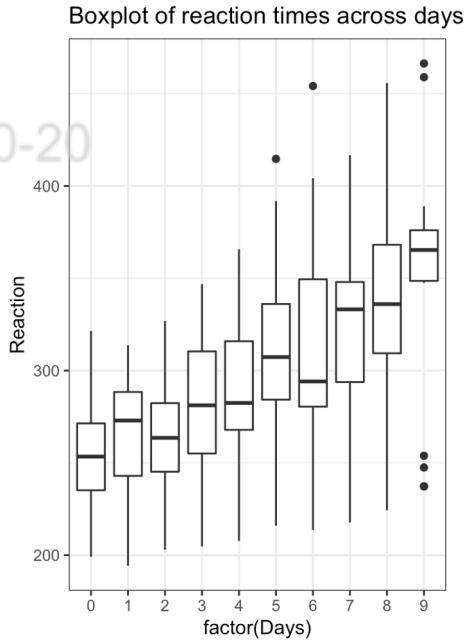
- **RM Anova**: **Separation** - avoiding variance under-estimation by fitting **group-level regressions per each unit of repetition**. Only works if N of observations is equal (because it cannot compare group-level regressions with different number of degrees of freedom)
- **GEE**: **Correlation** - avoiding variance under-estimation by adding **a correlation term between observations within one group**, "relaxing" the assumption of independence. Works well with missing data but does not allow associations between sources of variance and DV (may be of interest in longitudinal studies).
- **MM**: **nesting** - pulling out variance from the fixed effects by adding another kind of effect (the random) that allow individual level estimates of the coefficients

# Why individual level estimates of the coefficients?

**Working Example: Reaction Times In A Sleep Deprivation Study**

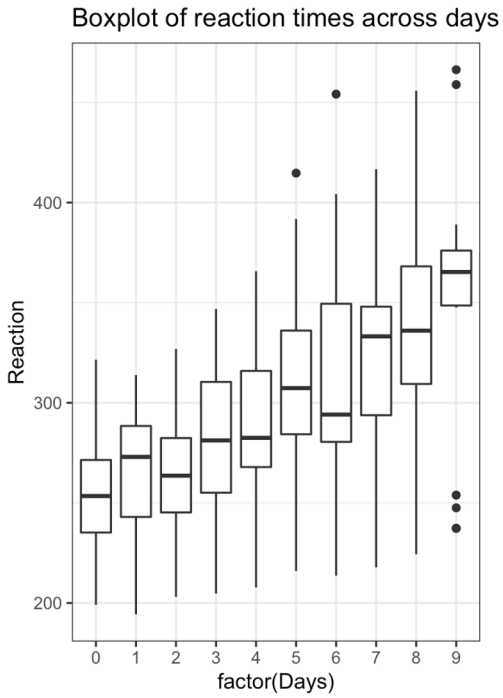


| Reaction | Days | Subject |
|---|---|---|
| 249.5600 | 0 | 308 |
| 258.7047 | 1 | 308 |
| 250.8006 | 2 | 308 |
| 321.4398 | 3 | 308 |

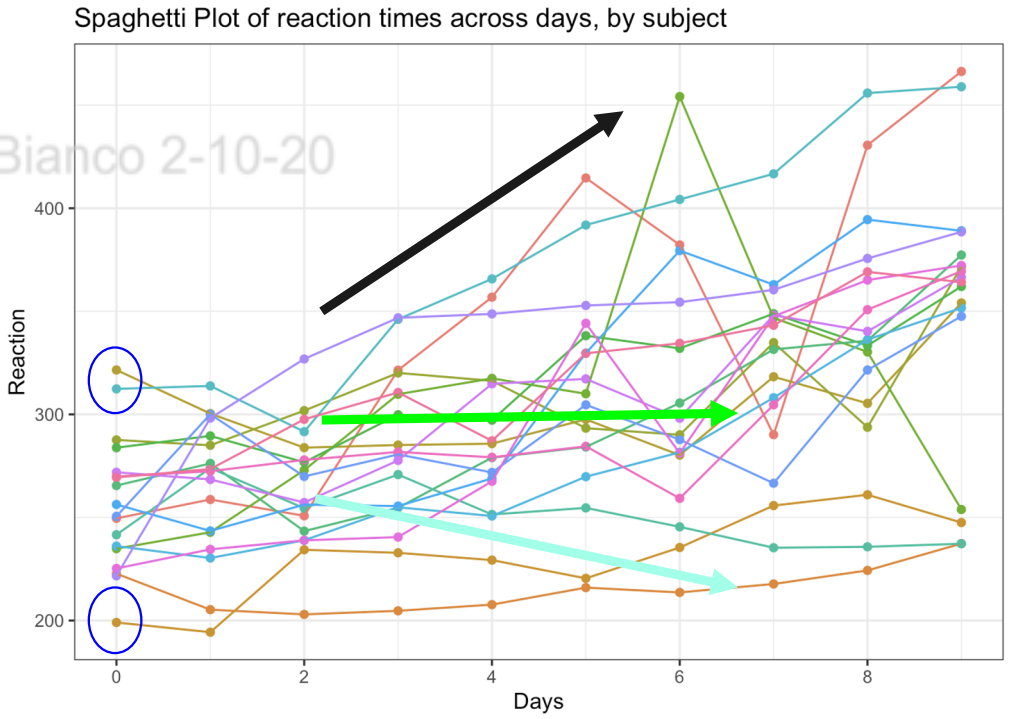



| Day | Avg Reaction |
|---|---|
| 0 | 257 |
| 1 | 264 |
| 2 | 265 |
| 3 | 283 |
| 4 | 289 |
| 5 | 309 |
| 6 | 312 |
| 7 | 319 |
| 8 | 337 |
| 9 | 351 |

# Why individual level estimates of the coefficients?





# The MM: Fixed Effects (B1)



- They represent **the independent variables** (IV), that affect the dependent variable (DV) in the same way for everybody (hence, *fixed*)
- The model estimates **the effect of the IV/fixed effect** (the *coefficient*), and its precision (the *standard error*).
- **The coefficient is on the same scale of the DV:** a very small estimate means that the IV has a very small effect on the DV. A negative estimate means that the IV decreases the DV.
- In other kinds of models, such as binomial models, where the DV takes the values of 0/1 (binary), the coefficient does not represent the average effect of the IV on the DV, but the

# The MM: Random Effects ($B_{0i}$ and $B_{1i}$)

- ~~**Random = Arbitrary**~~
- They specify different **starting points (intercepts)** and **degrees of variations (slopes)** for each subject.
- Why is it called random: I don't know really, but I made the analogy with the *Random Access Memory* (non-sequential)
- The random effect is a number describing **the deviation from the *fixed* coefficient** for each random effect.
- A random effect close to 0 means that, within that nesting, the deviation from the fixed coefficient is small.

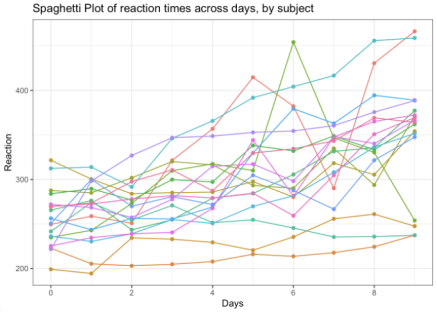
# The MM: Variance-Covariance Structure

- With repeated measures, variances are not homogeneus and observations co-vary (within subject, within time-point etc.)
- The easiest Variance-Covariance Structure is 'Unstructured'
- Many different types of structure can be implemented within a mixed model (diagonal, block diagonal, compound symmetry)
    - It introduces an assumption
    - It should be tested

# A Formula for the Working Example

**Working Example: Reaction Times In A Sleep Deprivation Study**

| Reaction | Days | Subject |
|---|---|---|
| 249.5600 | 0 | 308 |
| 258.7047 | 1 | 308 |
| 250.8006 | 2 | 308 |
| 321.4398 | 3 | 308 |




**Nesting: Day Between Subject**

**Formula:**



$$REACTION_i = B_0 + B_1*DAYS_i + B_{0i} + B_{1i}*DAYS_i + E_i$$

Fixed Effects

Random Effect

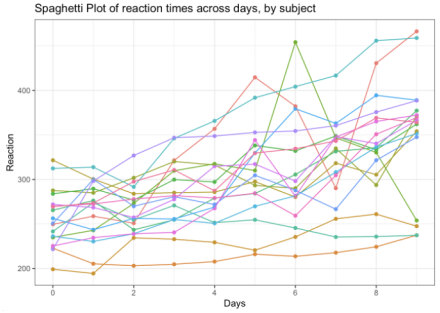
Varying **Intercept and Slope** per Level

B0 = Fixed Intercept
B1 = Fixed Coefficient Estimate of the IV "Days"
B0i = Varying Intercept (by subject)
B1i - Varying Slope (of the effect of the IV "Days" by subject)

# What are those Numbers? The Output

**Working Example: Reaction Times In A Sleep Deprivation Study**

| Reaction | Days | Subject |
|---|---|---|
| 249.5600 | 0 | 308 |
| 258.7047 | 1 | 308 |
| 250.8006 | 2 | 308 |
| 321.4398 | 3 | 308 |

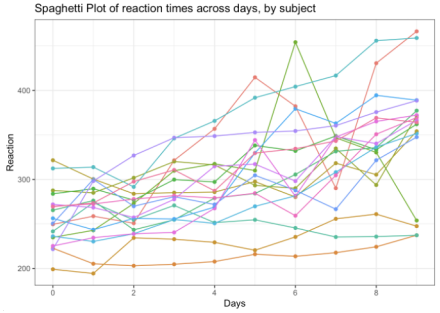



| Fixed Effect | Coefficient Estimate | Standard Error | DF | T-Value | P-Value | CI 2.5% | CI 97.5% |
|---|---|---|---|---|---|---|---|
| Intercept | 251.40 | 6.82 | 17 | 36.83 | <0.01 | 242.44 | 259.90 |
| Days | 10.46 | 1.54 | 17 | 6.77 | <0.01 | 8.70 | 12.82 |

# What are those Numbers? The Output

**Working Example: Reaction Times In A Sleep Deprivation Study**

| Reaction | Days | Subject |
|---|---|---|
| 249.5600 | 0 | 308 |
| 258.7047 | 1 | 308 |
| 250.8006 | 2 | 308 |
| 321.4398 | 3 | 308 |




| Random Effect | Variance | Standard deviation | Correlation |
|---|---|---|---|
| Subject (Intercept) | 612.10 | 24.74 | |
| Days (Slope) | 35.07 | 5.99 | 0.07 |
| | | | |
| Residual Variance | 654.94 | 25.59 | |

# How to Report

1. **The structure of the model** (what are you fixed effects, as well as the random effect: varying intercept and slope? Assumptions on the variance-covariance structure?)
2. **How you calculated p-values and confidence intervals** (and if the software does it for you, how is it doing it?)
3. **The p-value is less important** than the estimate and the standard error
4. **The estimate is an unstandardised effect size**
5. **The confidence interval** is another very good indicator of whether your model does really tell something or is only hyping the p-values

# Enriching the Output

- Comparisons between models
  - Fit a base model (without your IV of interest) and compare the model fit
  - Participant may differ on a certain covariate, but it might not improve your model fit afterall!
- Testing the Fixed Effects one against each other
  - Anova: tests the significance of one factor while controlling for the level of other factors
  - Linear Contrasts: run pairwise significance tests between levels of one factor
- Plotting
  - The effects as lines
  - Confidence Intervals
- Testing the Assumptions
  - Normality of Residuals

# Cool Stuff that you did not know your Model could do

Individual Coefficients

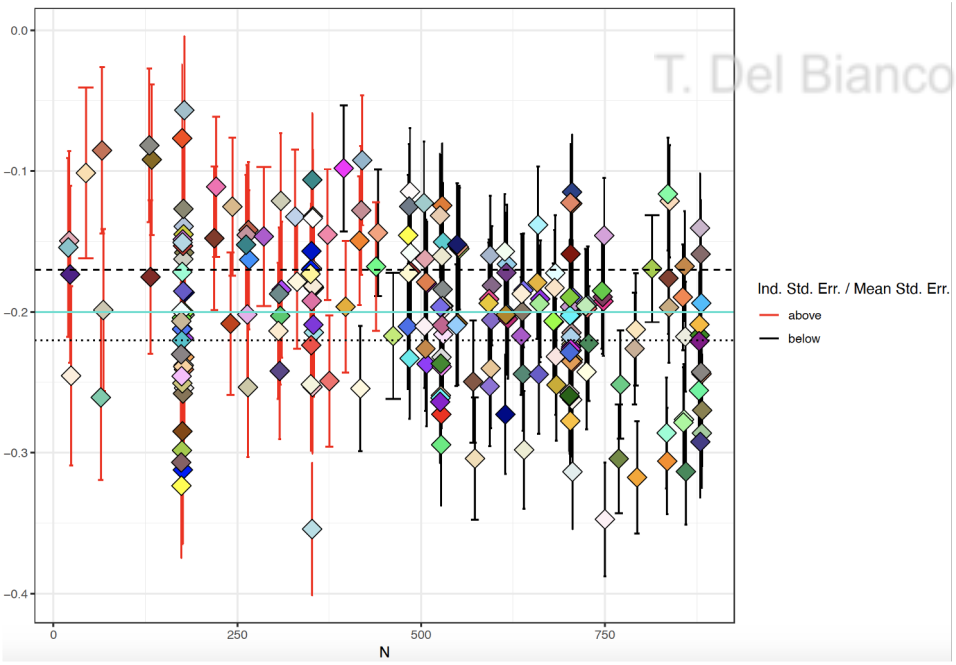
```
## $Subject
##     (Intercept)       Days
## 308    253.6637 19.6662617
## 309    211.0064  1.8476053
## 310    212.4447  5.0184295
## 330    275.0957  5.6529356
## 331    273.6654  7.3973743
## 332    260.4447 10.1951090
## 333    268.2456 10.2436499
## 334    244.1725 11.5418676
## 335    251.0714 -0.2848792
## 337    286.2956 19.0955511
## 349    226.1949 11.6407181
## 350    238.3351 17.0815038
```



Random Effects

```
## $Subject
##     (Intercept)        Days
## 308   2.2585509   9.1989758
## 309 -40.3987381  -8.6196806
## 310 -38.9604090  -5.4488565
## 330  23.6906196  -4.8143503
## 331  22.2603126  -3.0699116
## 332   9.0395679  -0.2721770
## 333  16.8405086  -0.2236361
## 334  -7.2326151   1.0745816
## 335  -0.3336684 -10.7521652
## 337  34.8904868   8.6282652
```

# Cool Stuff that you did not know your Model could do




Y = Individual Coefficients

X = N of observations per participant

Error Bar = summated variance of the fixed and the random effects

# Additional Resources

- A great book https://g.co/kgs/hYqLqr
- A great couple of tutorials http://www.bodowinter.com/resources.html

# Thanks for Listening!