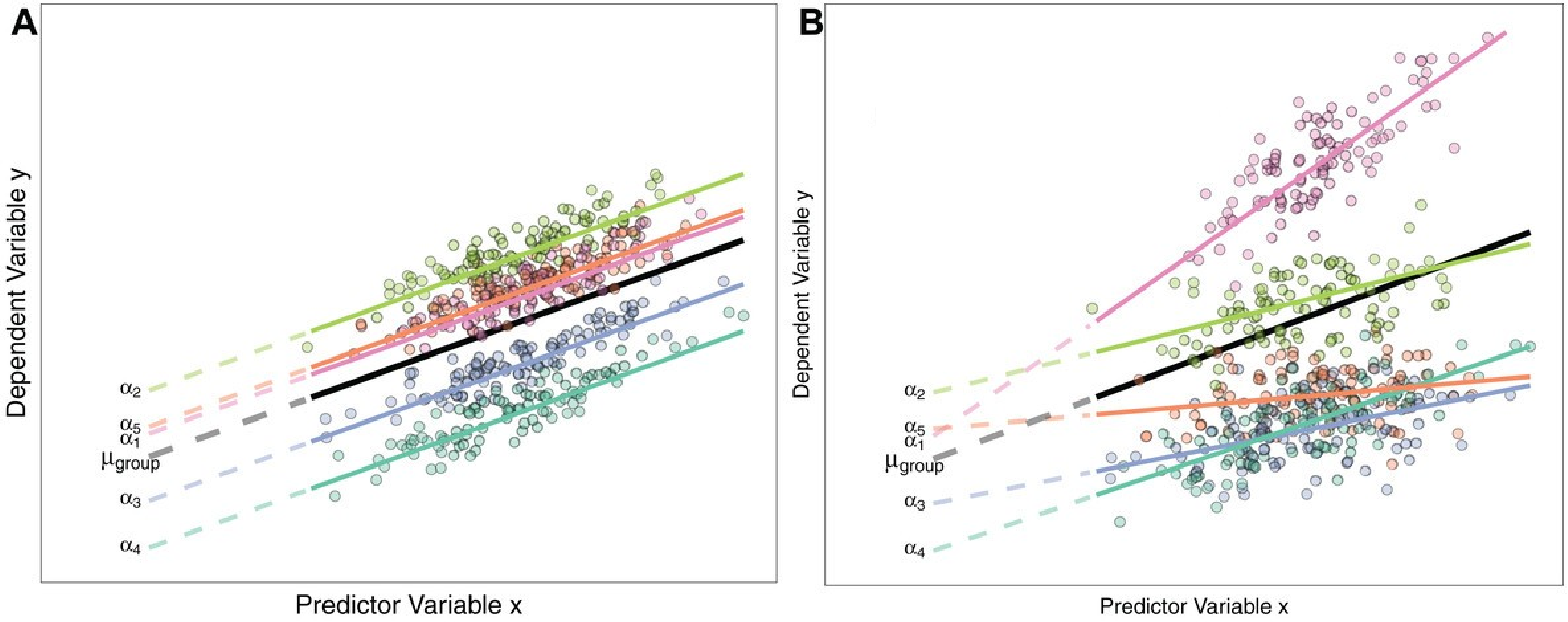
Introduction to:

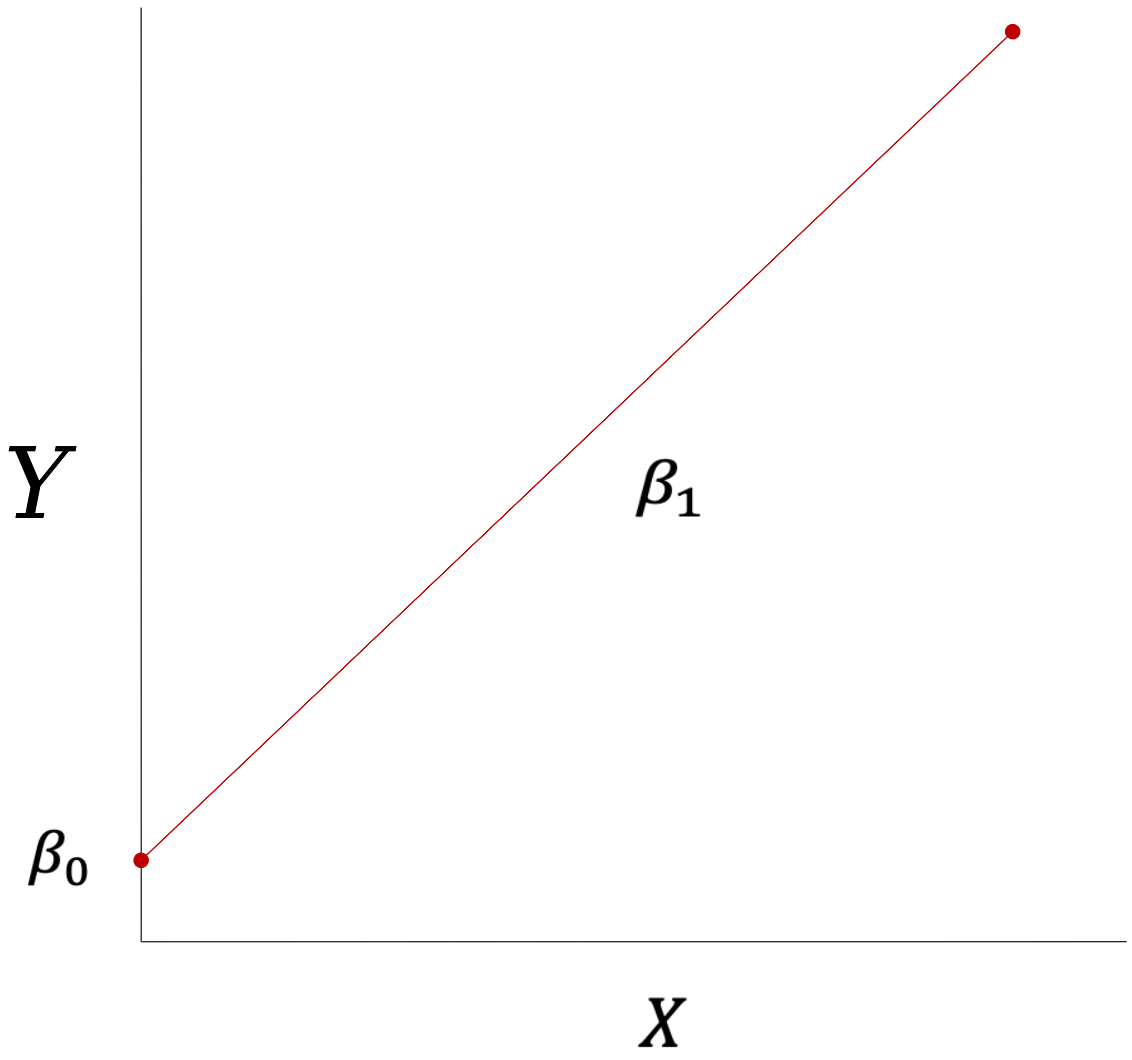
Linear Mixed Effects models



Linear mixed models
Multilevel models
Hierarchical linear models

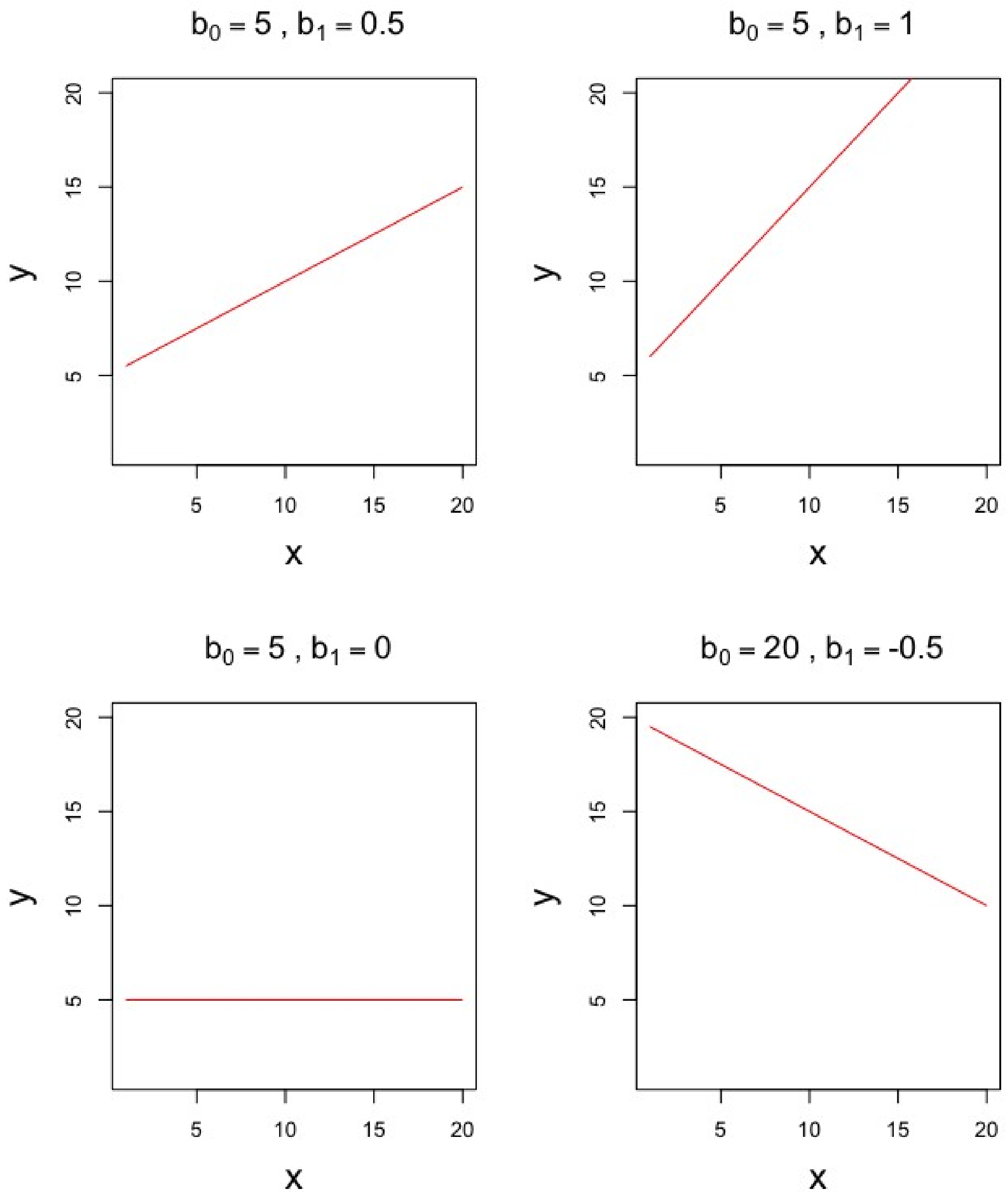
Linear mixed models

$$Y = \beta_0 + \beta_1 X$$



Linear mixed models

$$Y = \beta_0 + \beta_1 X$$



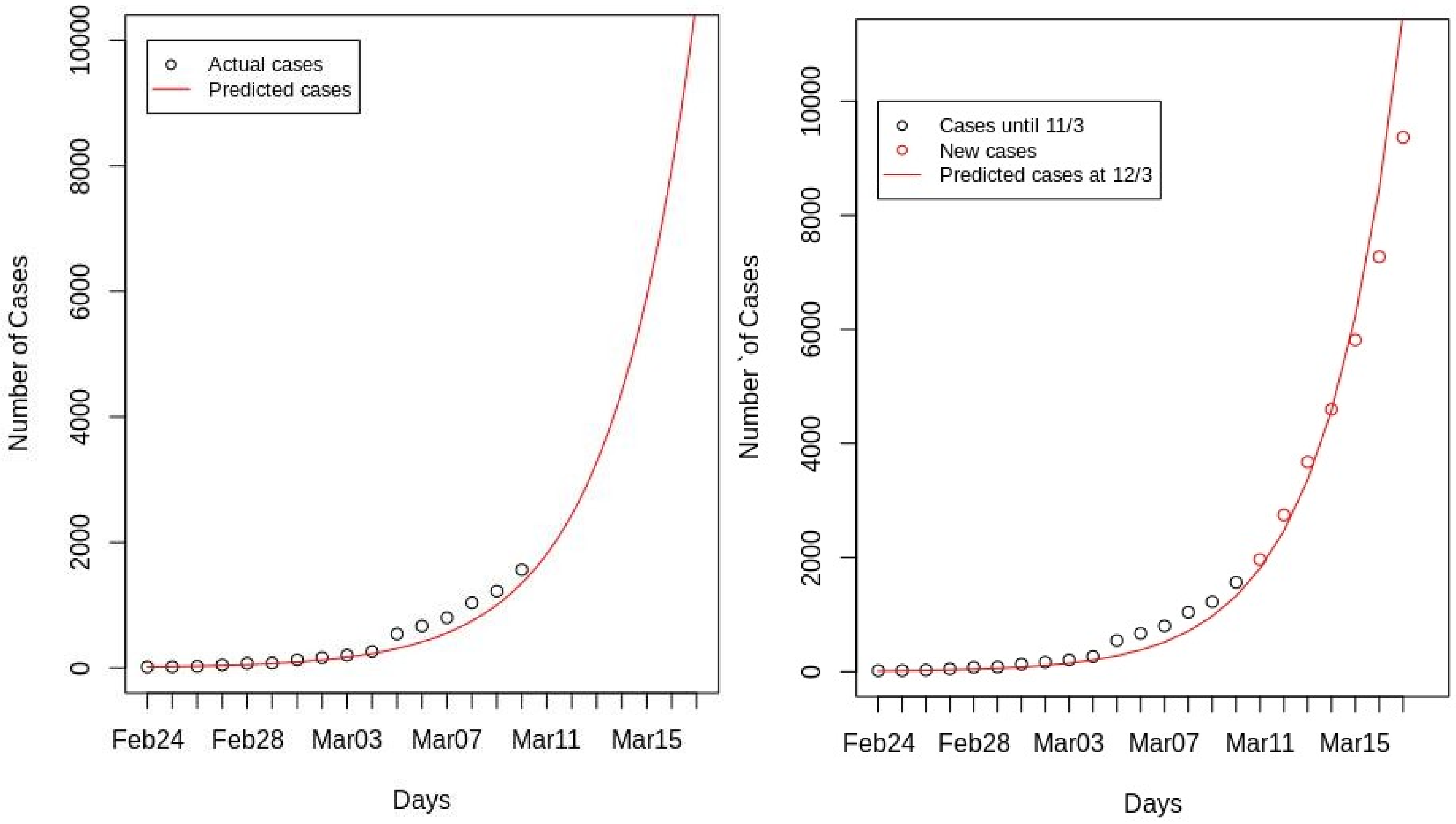
Linear mixed models

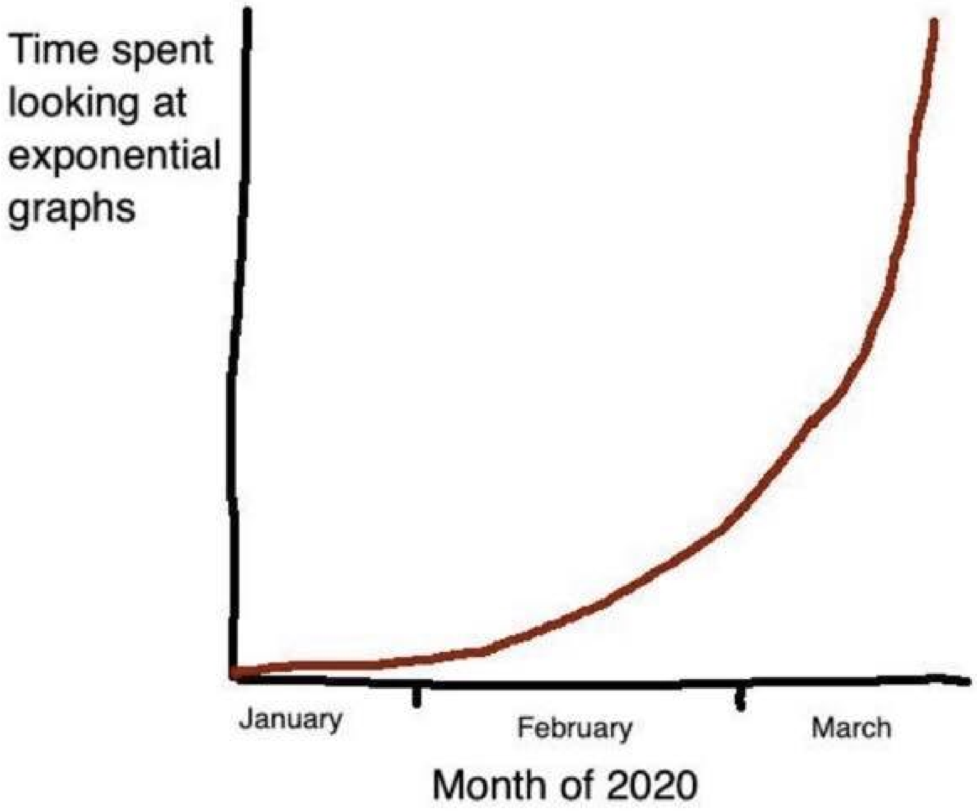
It is a linear model because the parameters combine additively; it is the model that needs to be linear, not necessarily the relationship(s)

$$Y = ae^{\beta x}$$

$$\ln Y = \ln \alpha + \beta x$$

Covid-19 cases in Germany

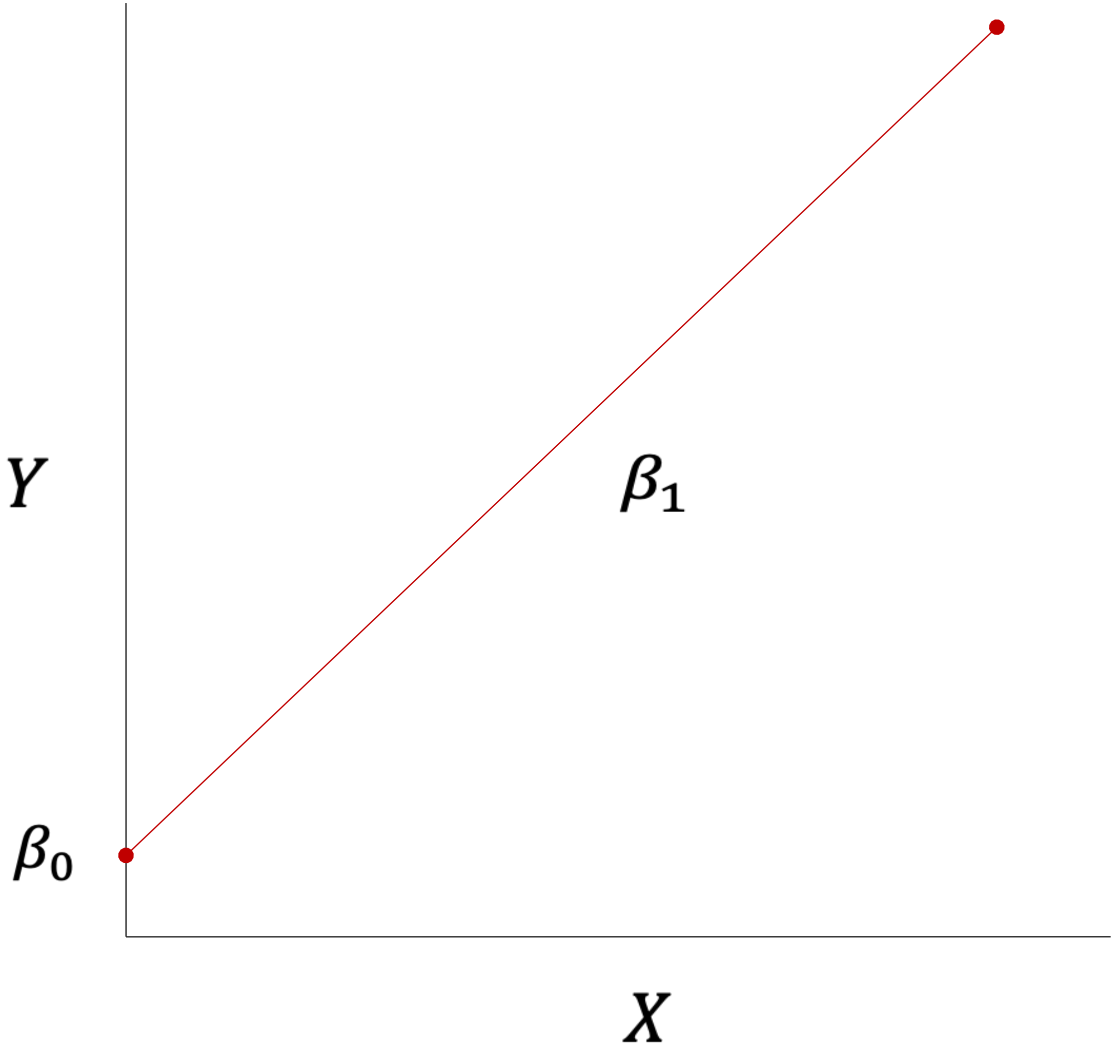



Time spent looking at exponential graphs
January
February
March
Month of 2020

$$y_j = \beta_0 + \beta_1 x_j + \epsilon_j$$

Linear mixed models

Fixed effects: we assume that the model holds true across the entire sample and that for every case of data (participant) in the sample we can predict a score using the same values of the slope and intercept, plus some random error that represents all factors that might influence the dependent variable other than x.

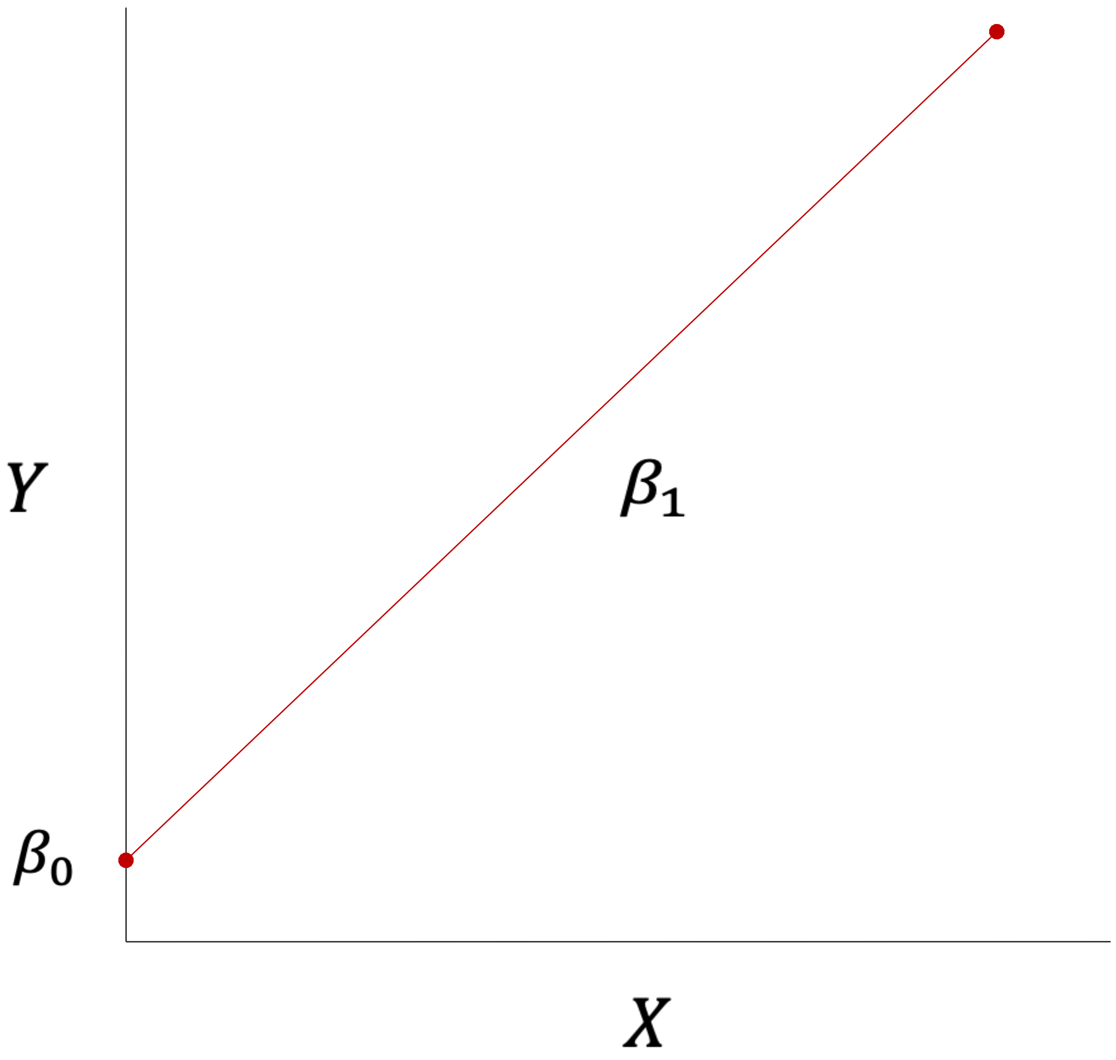


$$y_j = \beta_0 + \beta_1 x_j + \epsilon_j$$

Linear mixed models

In linear models, the parameters β_0 and β_1 are estimated through OLS.

It minimizes the sum of the squared differences between the observed values of *y* and the model-predicted values of *y* across the entire sample.



Lmmworkshop.Rmd

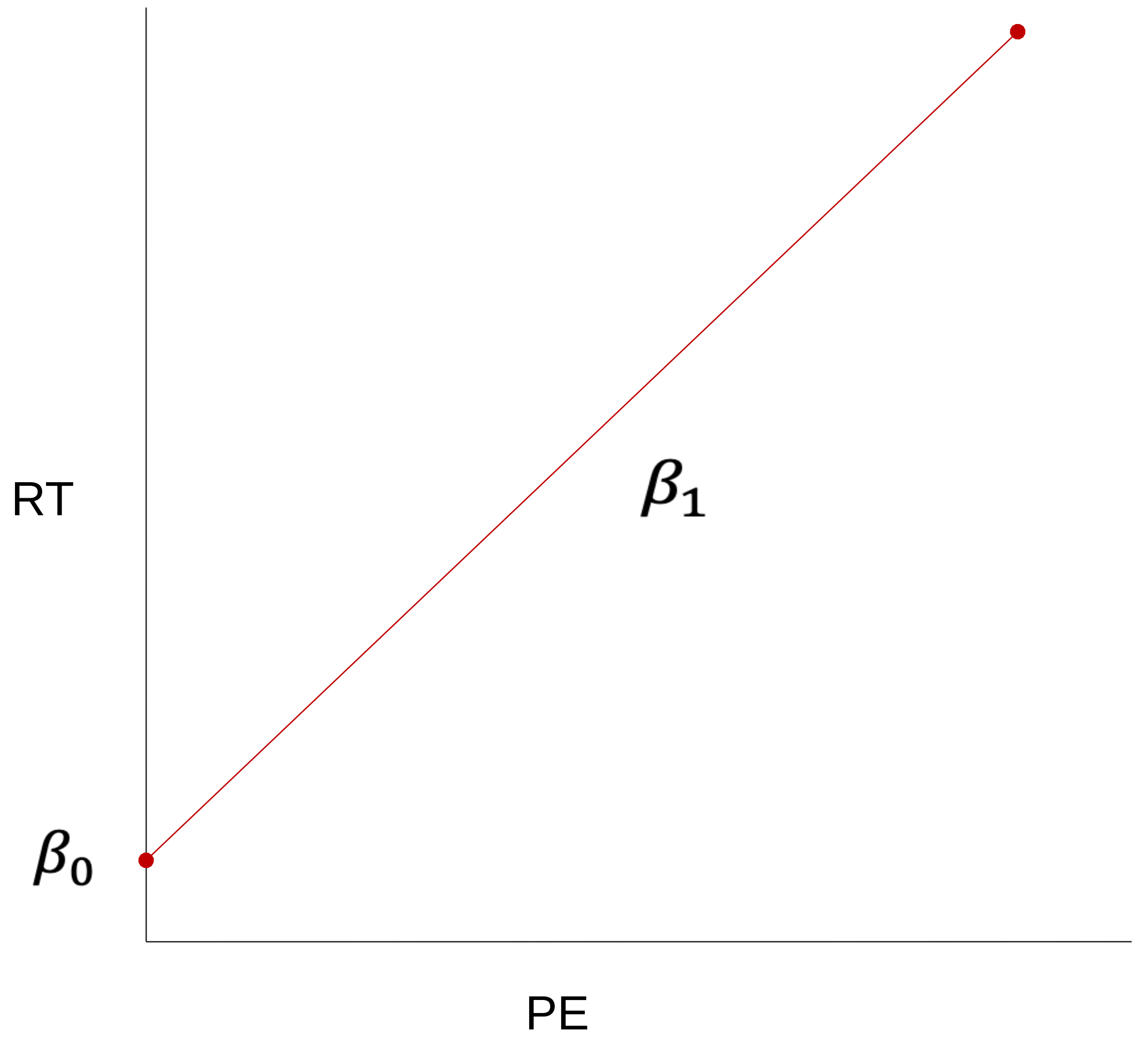
```{r simulate Data}
```

Linear mixed models

What if we have variables that vary within-participants?

subj_id	item_id	PE	RT
1	1	-2.24718275	3.499112
1	2	0.59722872	3.553017
1	3	2.72881185	3.692639
1	4	-1.72184306	3.295054
1	5	1.07061702	3.166503
1	6	1.01296702	3.070292
1	7	0.84770910	3.440259
1	8	-0.19847366	3.246503
1	9	-0.02643587	3.112401
1	10	0.72538218	3.548040
1	11	0.51826827	3.457657
1	12	-0.72260728	3.783603
1	13	-0.73772957	3.007351
1	14	-0.14433796	3.220507
1	15	1.15225230	3.423374
1	16	0.79673807	3.004323
1	17	-1.93117571	3.193946
1	18	-1.61749030	3.083845

$$y_j = \beta_0 + \beta_1 x_j + \epsilon_j$$



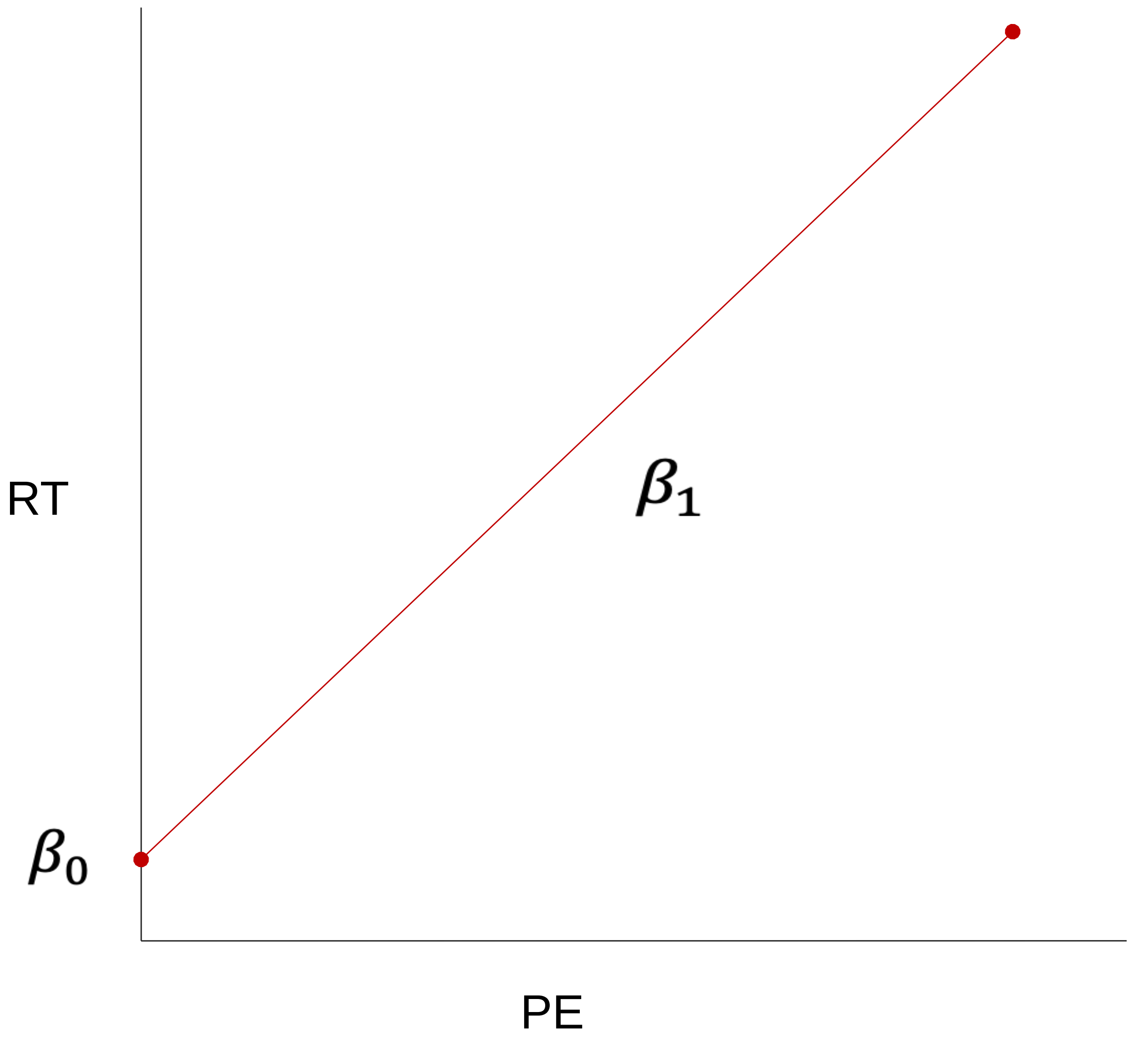
$$y_j = \beta_0 + \beta_1 x_j + \epsilon_j$$

Linear mixed models

Solution 1: We could fit a OLS regression to the entire
Dataset, without aggregating.

What is the problem with that?

Exactly: it violates the assumptions of indipendence, because participants provide more than one data point. Therefore, their data are correlated (there are clusters in the data).



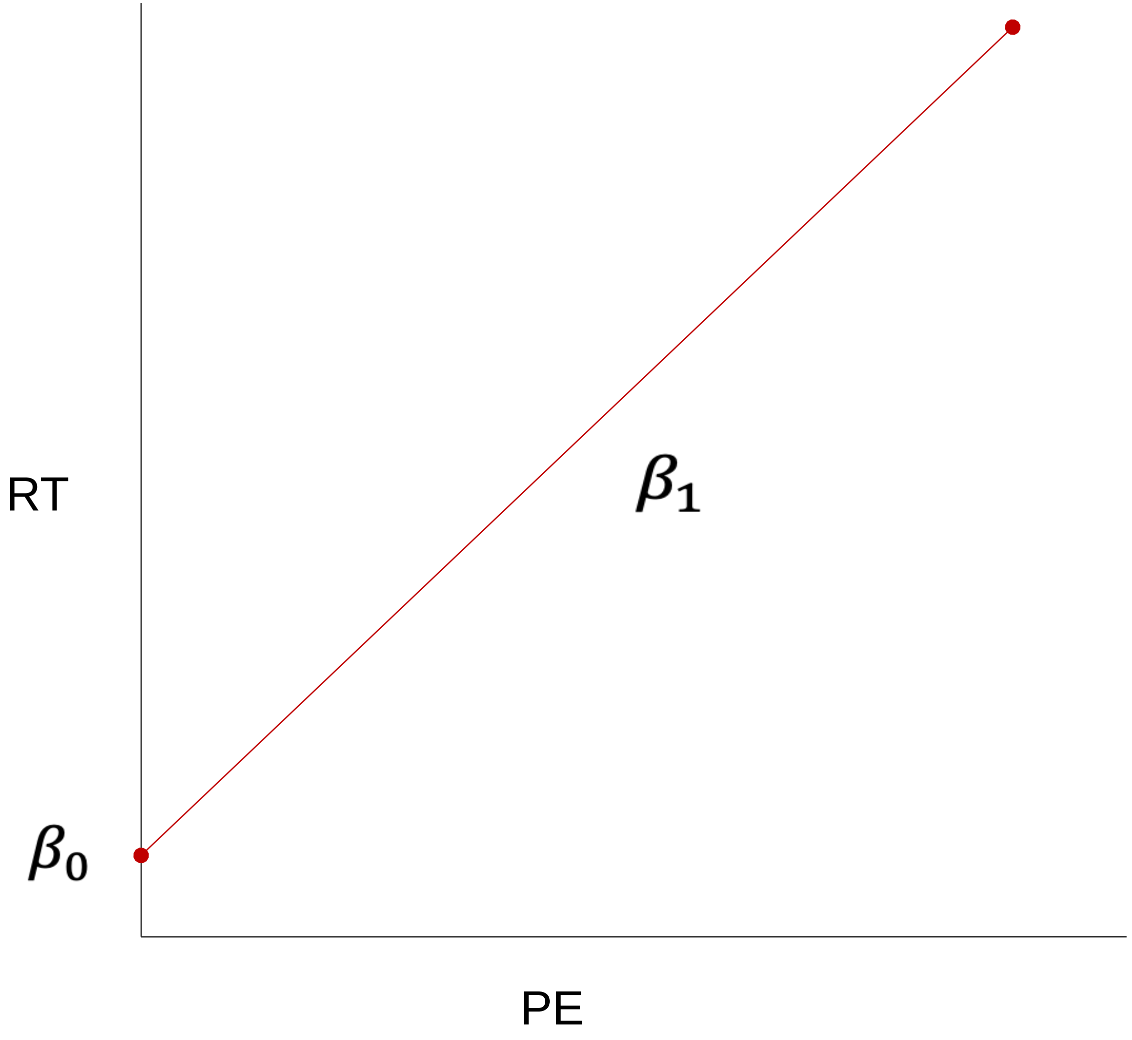
Linear mixed models

$$y_j = \beta_0 + \beta_1 x_j + \epsilon_j$$

Solution 2: We could aggregate data at the participant level and run OLS regression.

Lmmworkshop.Rmd

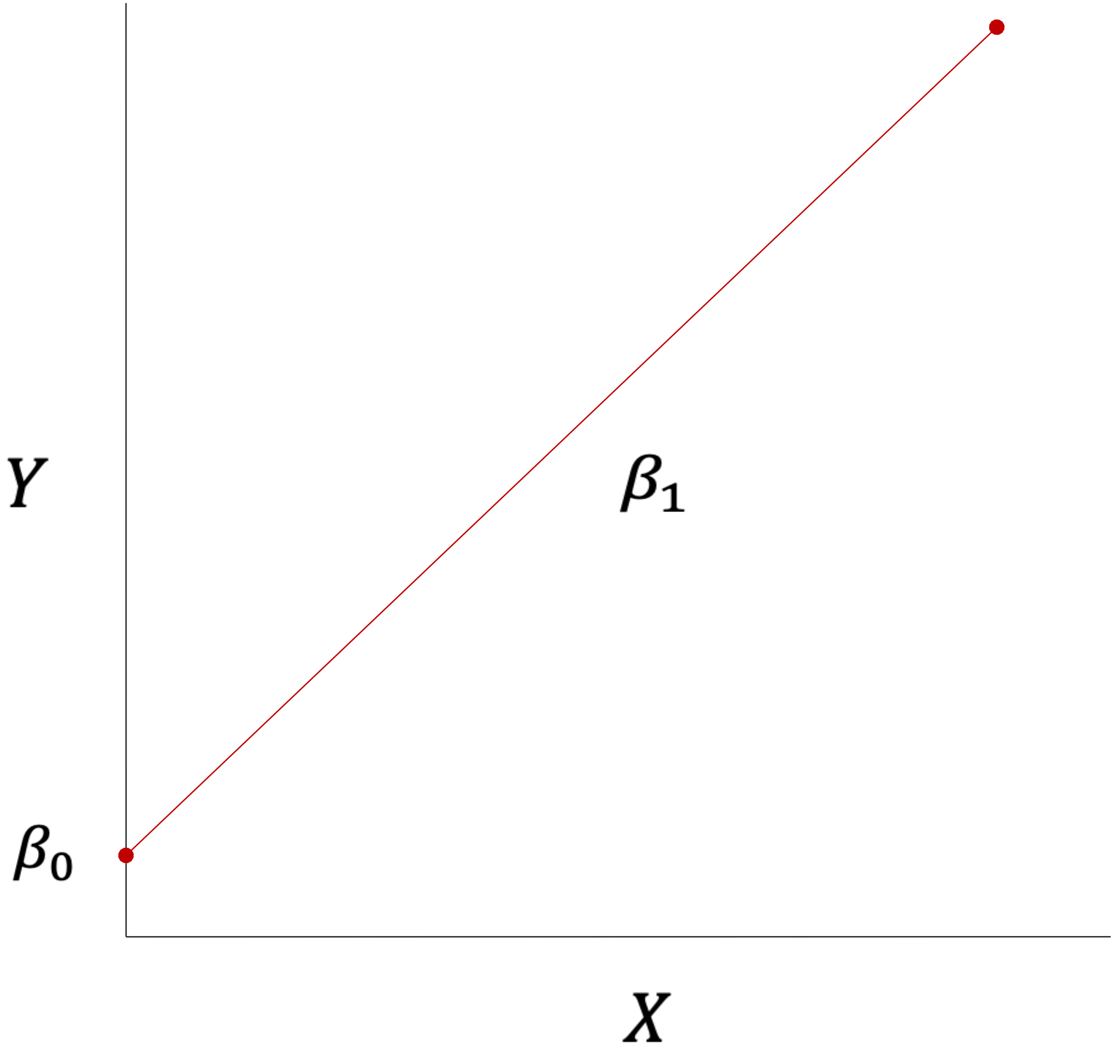
```{r aggregate }
```



$$y_j = \beta_0 + \beta_1 x_j + \epsilon_j$$

Linear mixed models

```
linearmodel<-lm(outcome~predictor,data=df)
```



Linear mixed models

summary(linearmodel)

```
Call:
lm(formula = RT ~ PE, data = df_agg)

Residuals:
   Min     1Q Median     3Q    Max
-4.192 -1.133 -0.148  1.186  3.372

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    0.852      6.860   0.124    0.902
PE             6.388     35.001   0.183    0.856

Residual standard error: 1.778 on 28 degrees of freedom
Multiple R-squared:  0.001188, Adjusted R-squared:  -0.03448
F-statistic: 0.03331 on 1 and 28 DF,  p-value: 0.8565
```

$$y_j = \beta_0 + \beta_1 x_j + \epsilon_j$$

Linear mixed models

summary(linearmodel)

```
Call:
lm(formula = RT ~ PE, data = df_agg)

Residuals:
   Min     1Q Median     3Q    Max
-4.192 -1.133 -0.148  1.186  3.372

Coefficients:
            Estimate Std. Error  t value  Pr(>|t|)
(Intercept)    0.852      6.860    0.124    0.902
PE             6.388     35.001    0.183    0.856

Residual standard error: 1.778 on 28 degrees of freedom
Multiple R-squared:  0.001188, Adjusted R-squared:  -0.03448
F-statistic: 0.03331 on 1 and 28 DF,  p-value: 0.8565
```

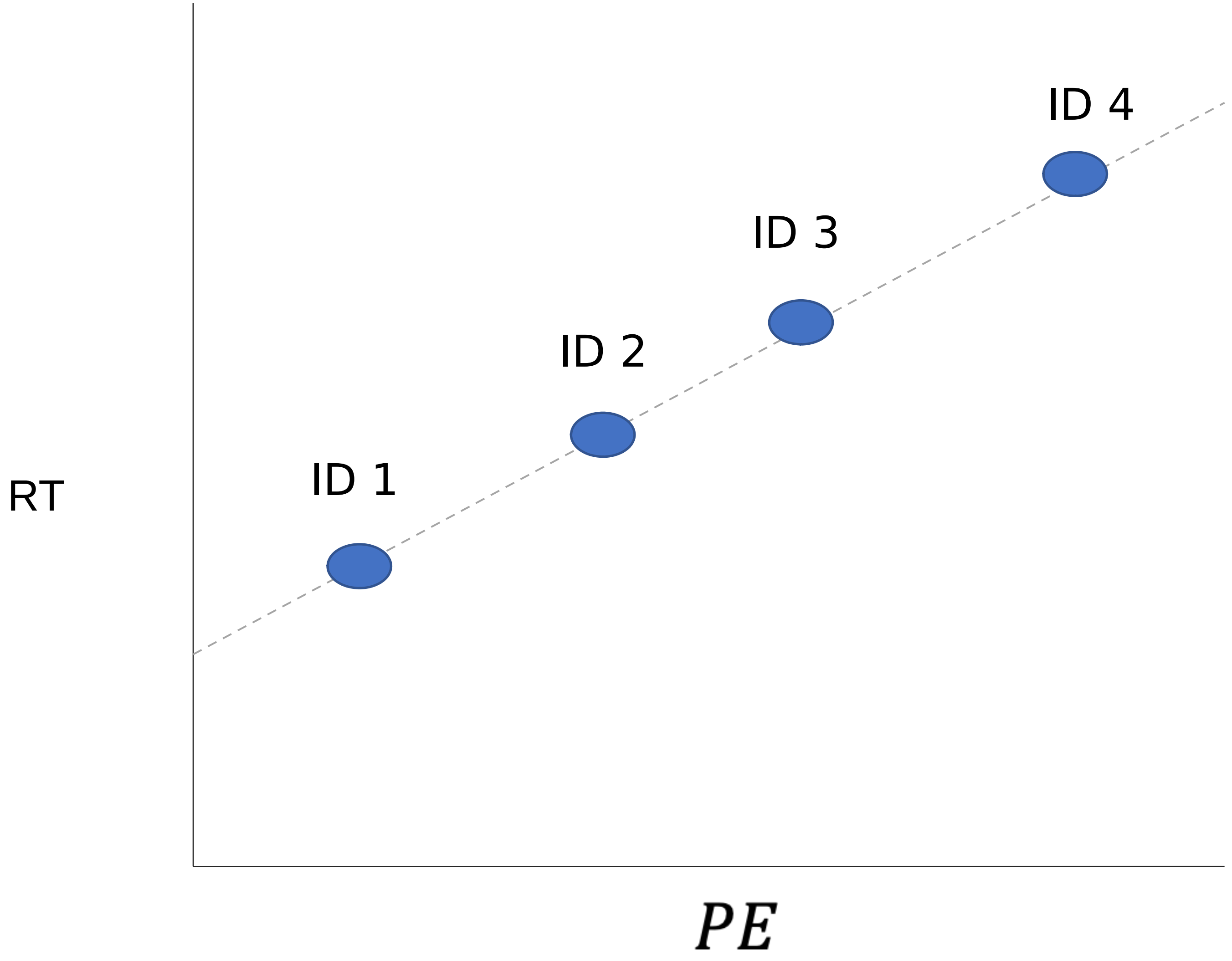
$$y_j = \beta_0 + \beta_1 x_j + \epsilon_j$$

Linear mixed models

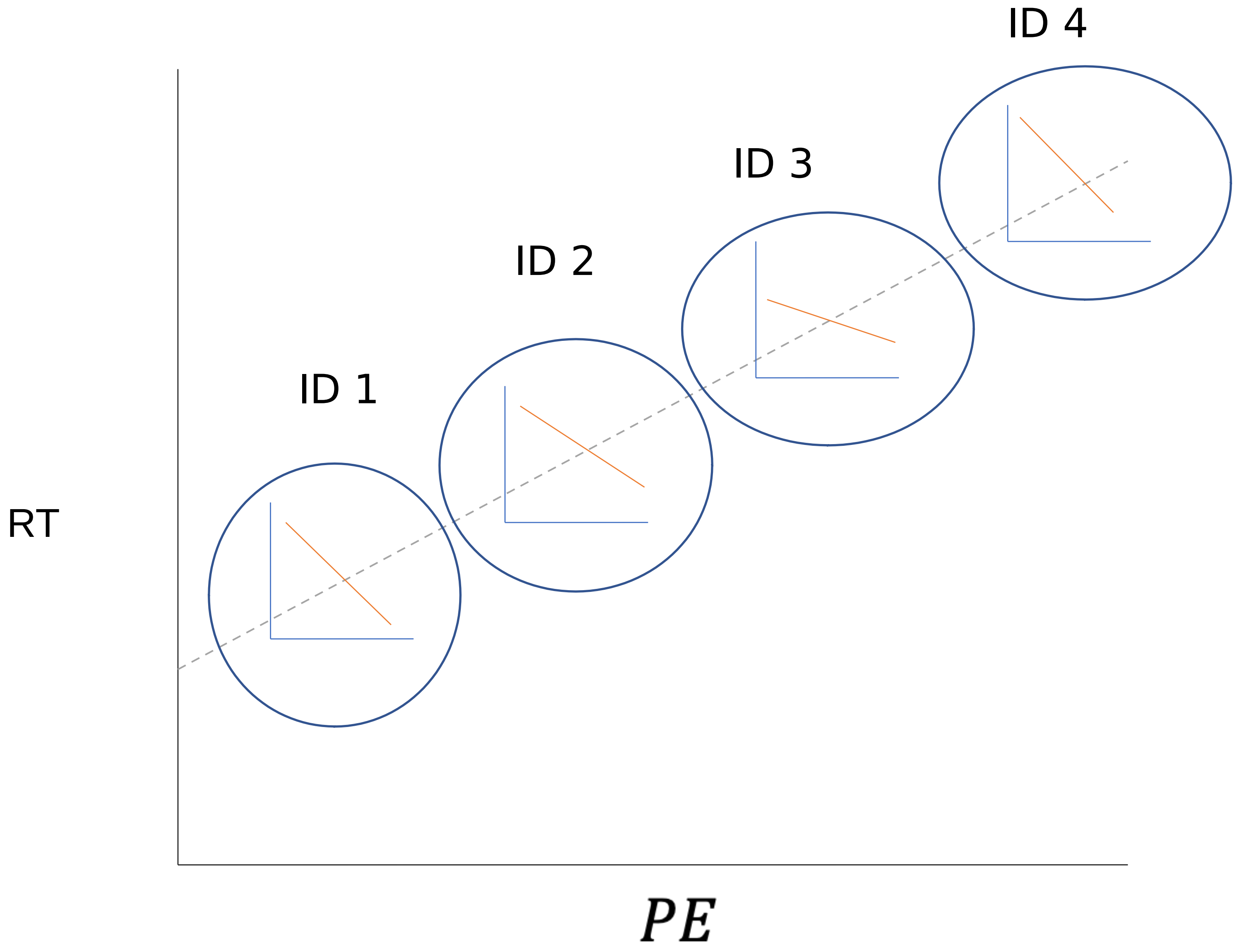
Summary.aov(linearmodel)

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
PE	1	0.11	0.1053	0.033	0.856
Residuals	28	88.51	3.1612		

What is the problem with that?



Between-participant level:
Participants with higher PE on average have slower RT

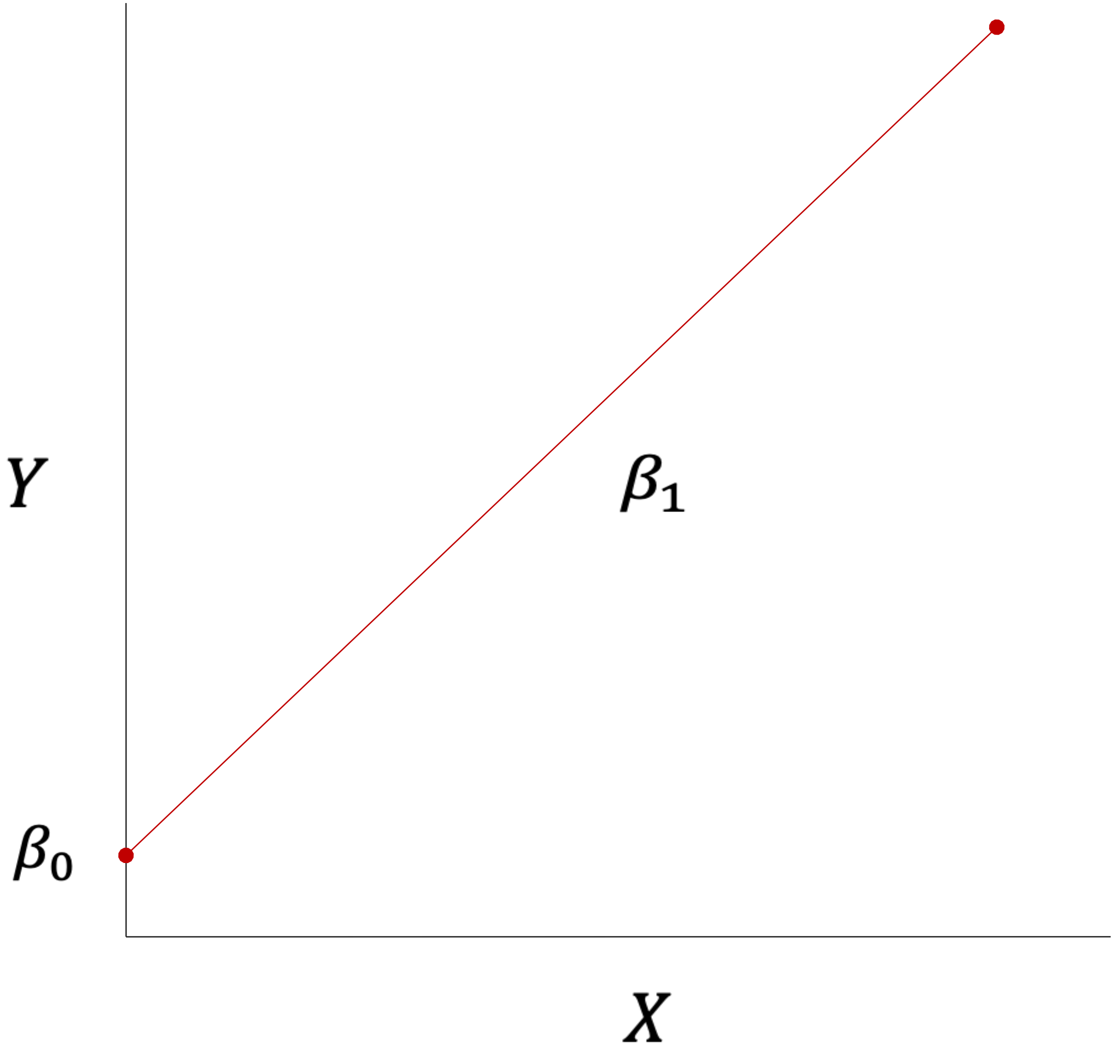


Within-participant level:
When participants experienced higher PE their RT are faster

Linear mixed models

$$y_j = \beta_0 + \beta_1 x_j + \epsilon_j$$

So what do we do?

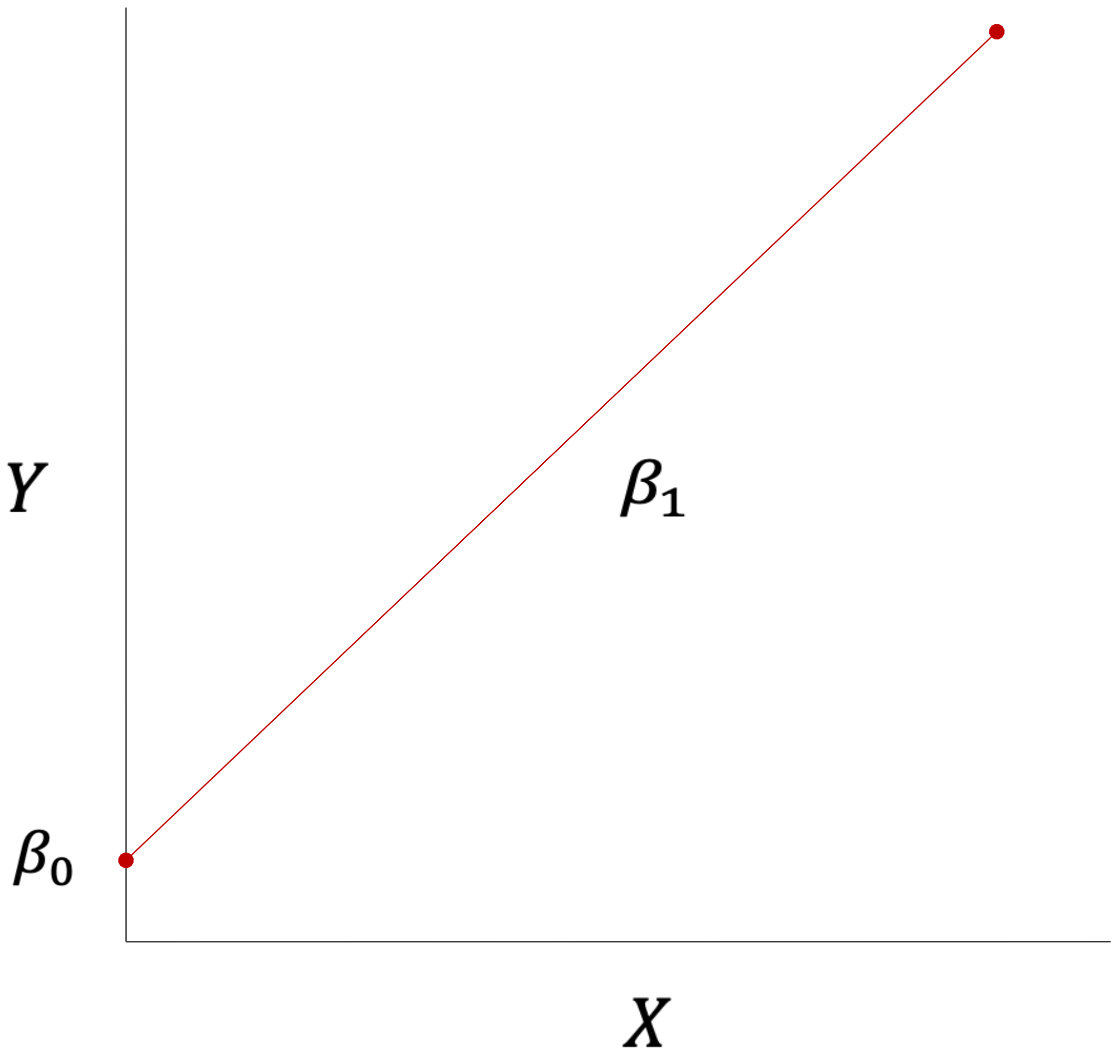


$$y_j = \beta_0 + \beta_1 x_j + \epsilon_j$$

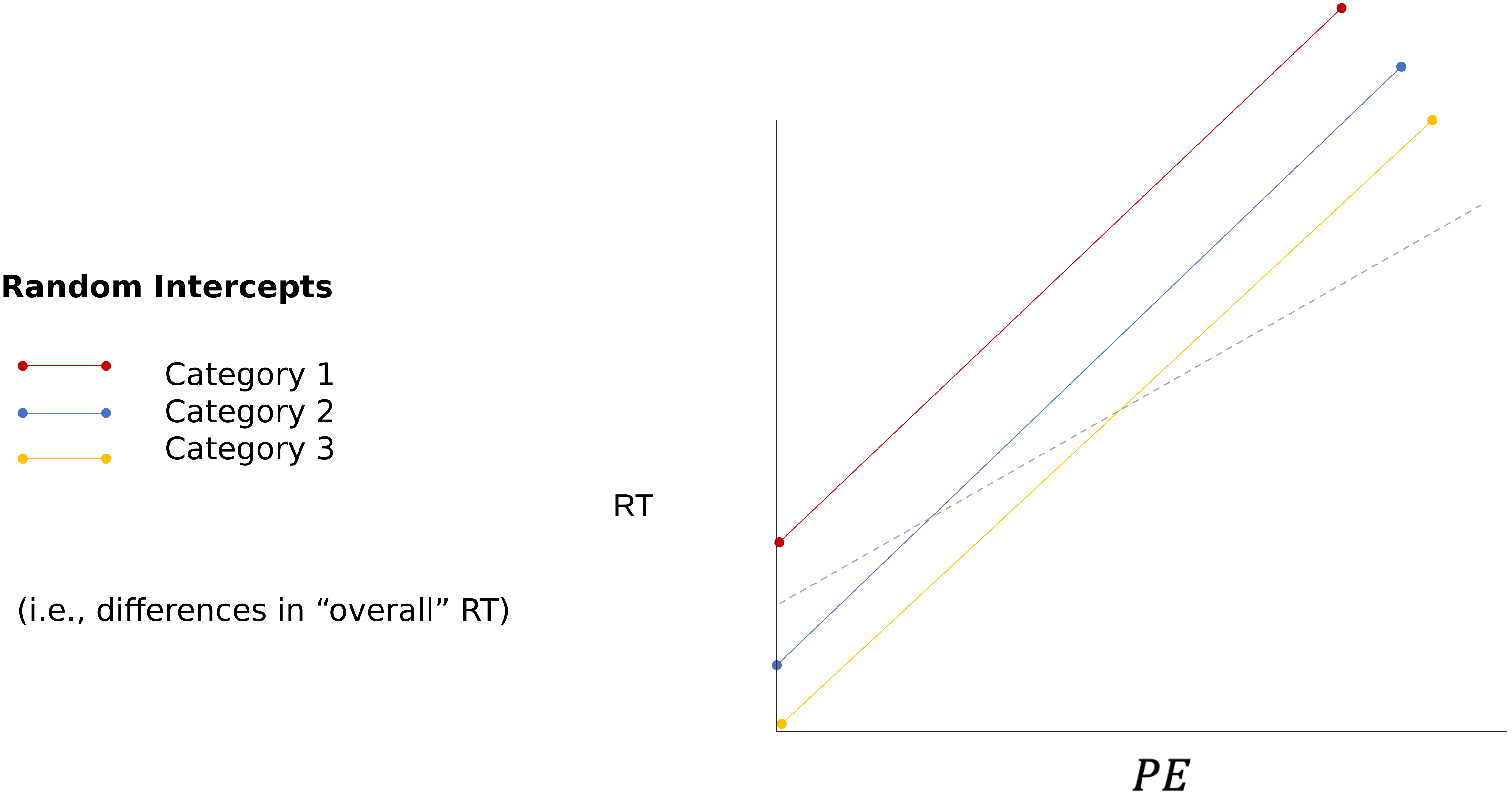
Linear mixed models

Random effects: we say that a parameter is random then we assume not that it is a
fixed value, but that its value can vary.

That parameter is also random because we treat it is as being randomly selected from the larger population.

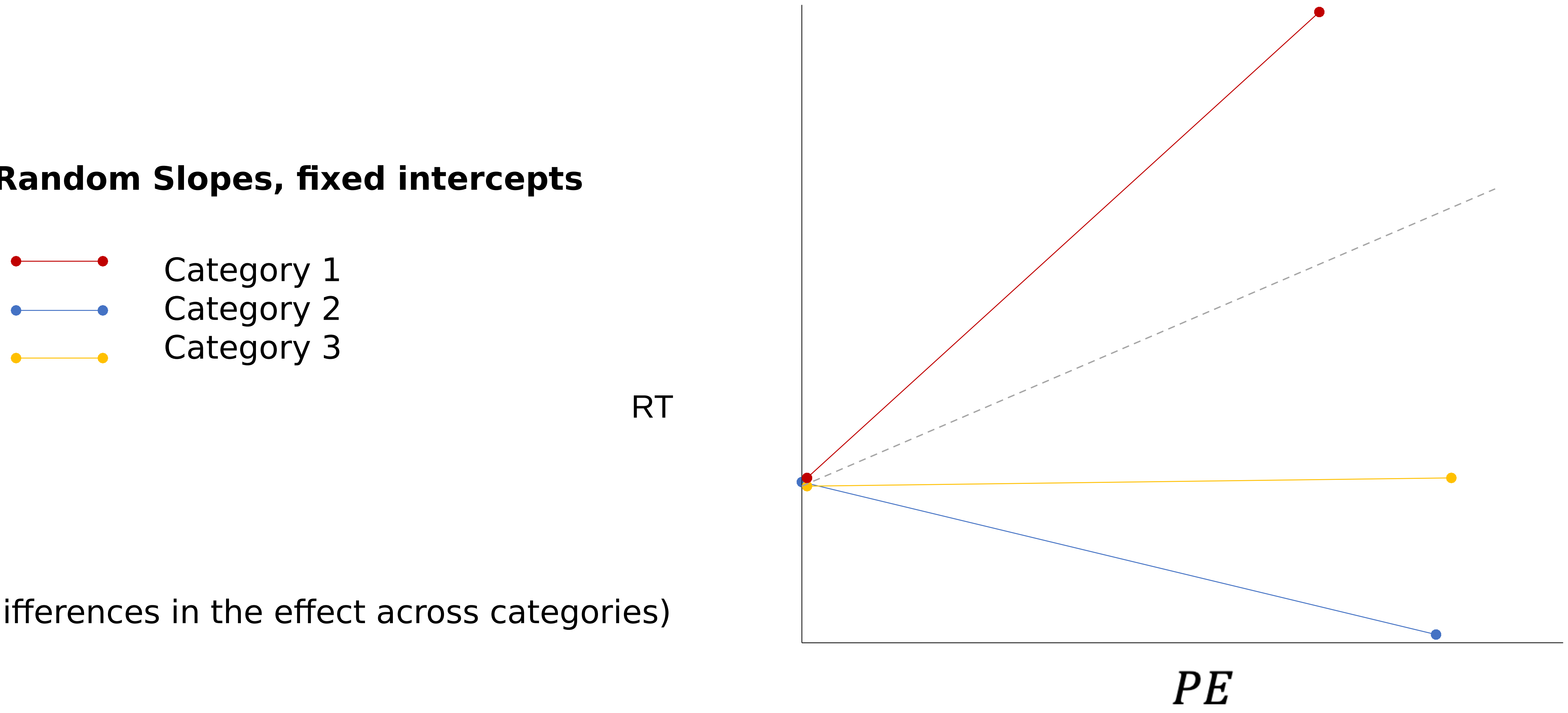


Random Intercepts



(i.e., differences in “overall” RT)

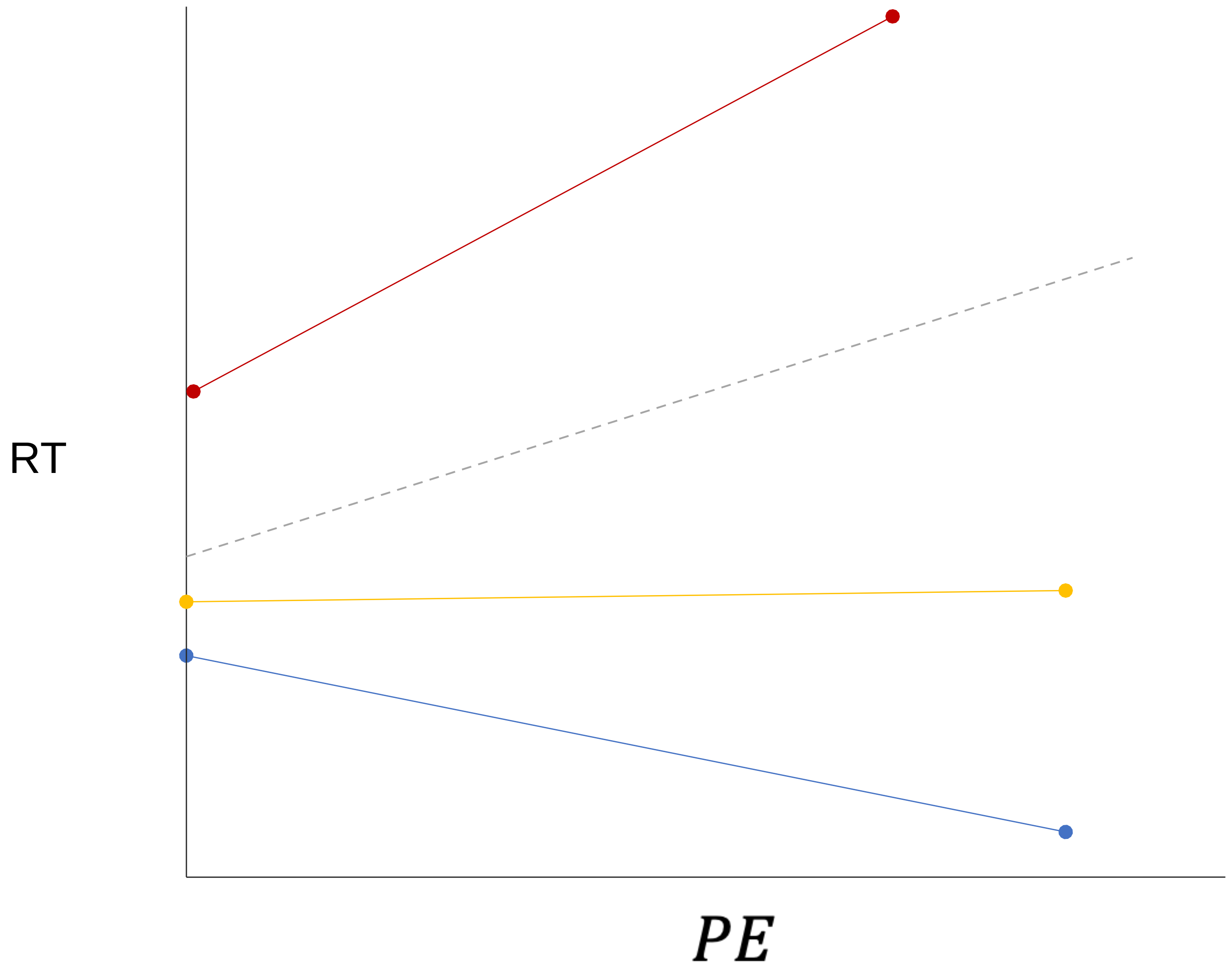
Random Slopes, fixed intercepts



i.e., differences in the effect across categories)

Random Intercepts and Slopes

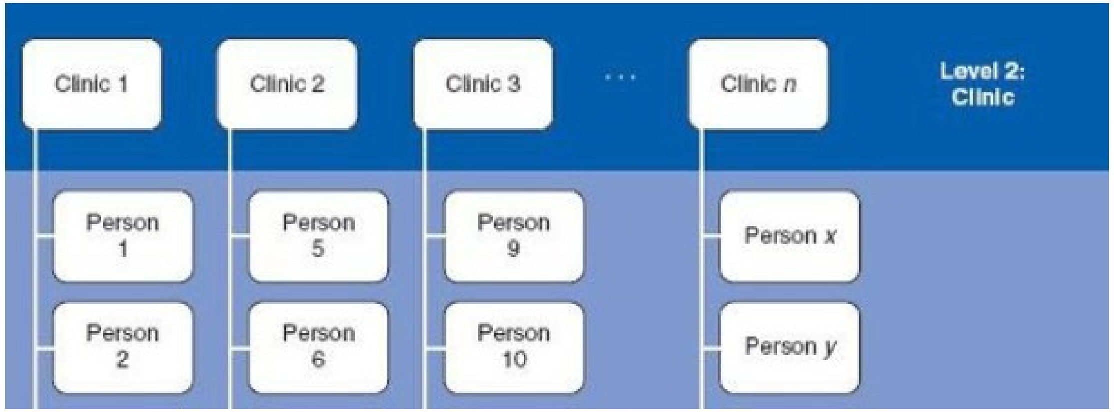
Category 1
Category 2
Category 3



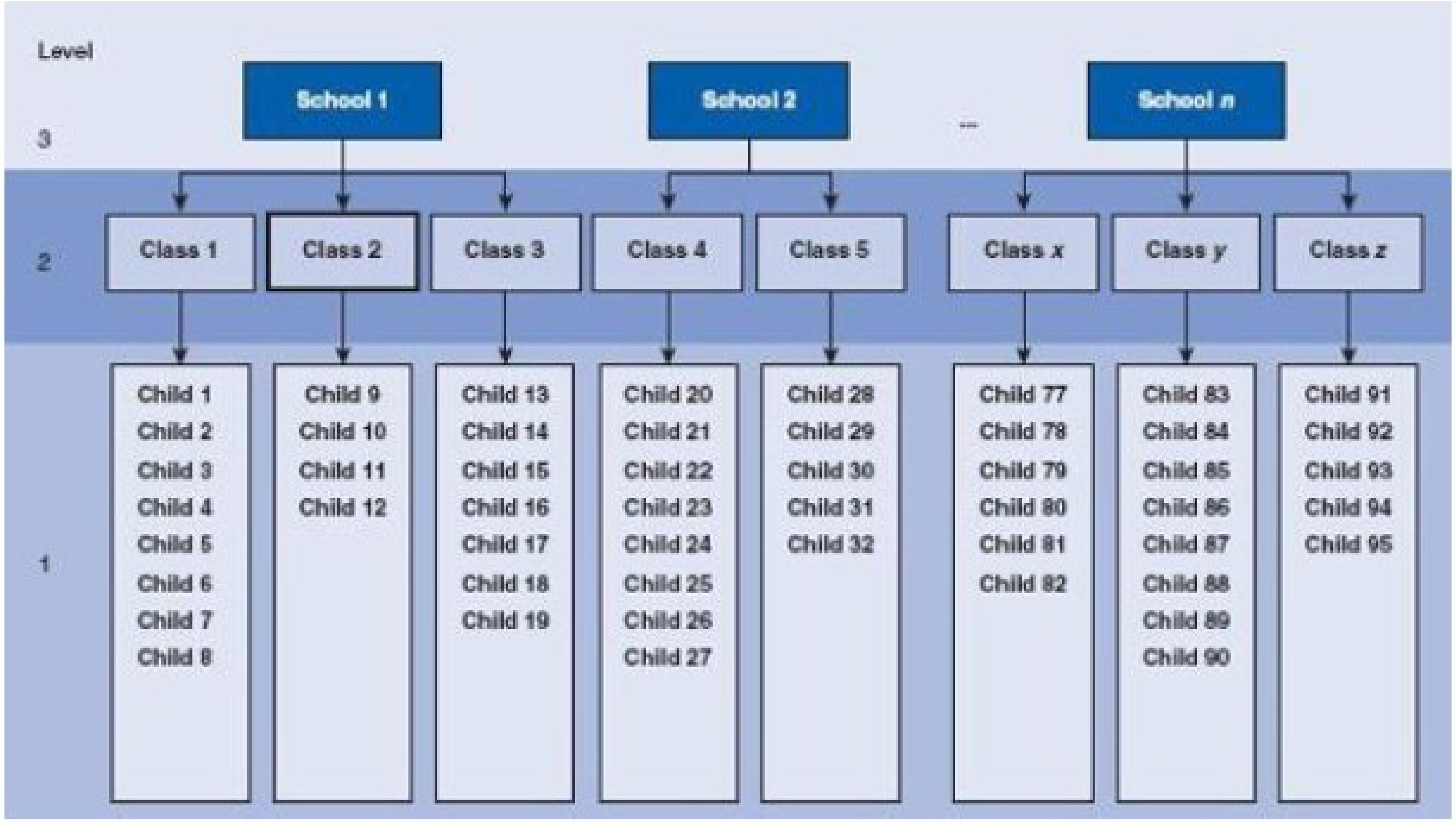
(i.e., differences in the size of the effect across conditions)

Including random intercepts and slopes allows to control for random variations among items (sampling units)

Linear mixed models
Multilevel models
Hierarchical linear mod



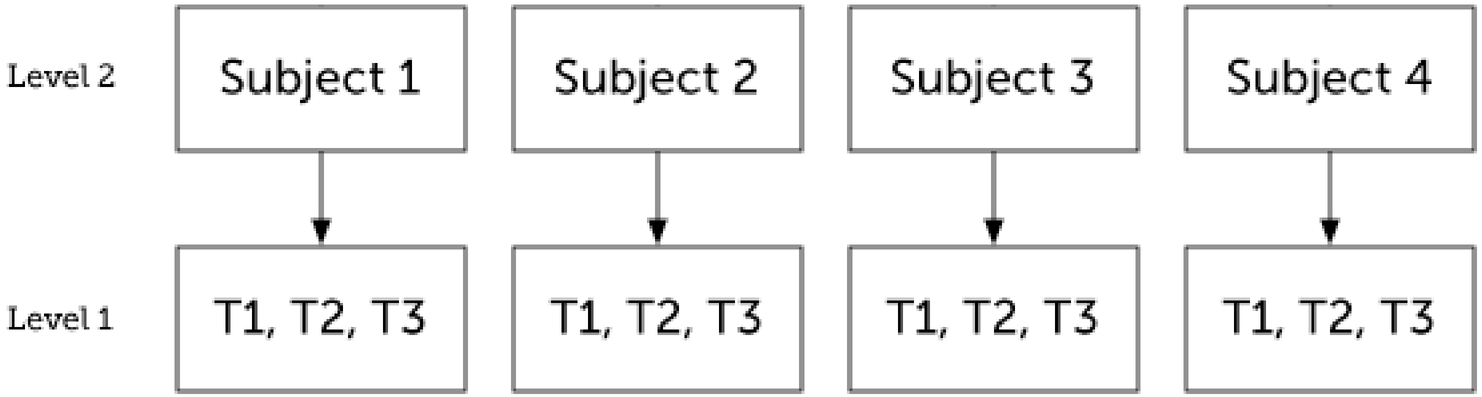
Nested data

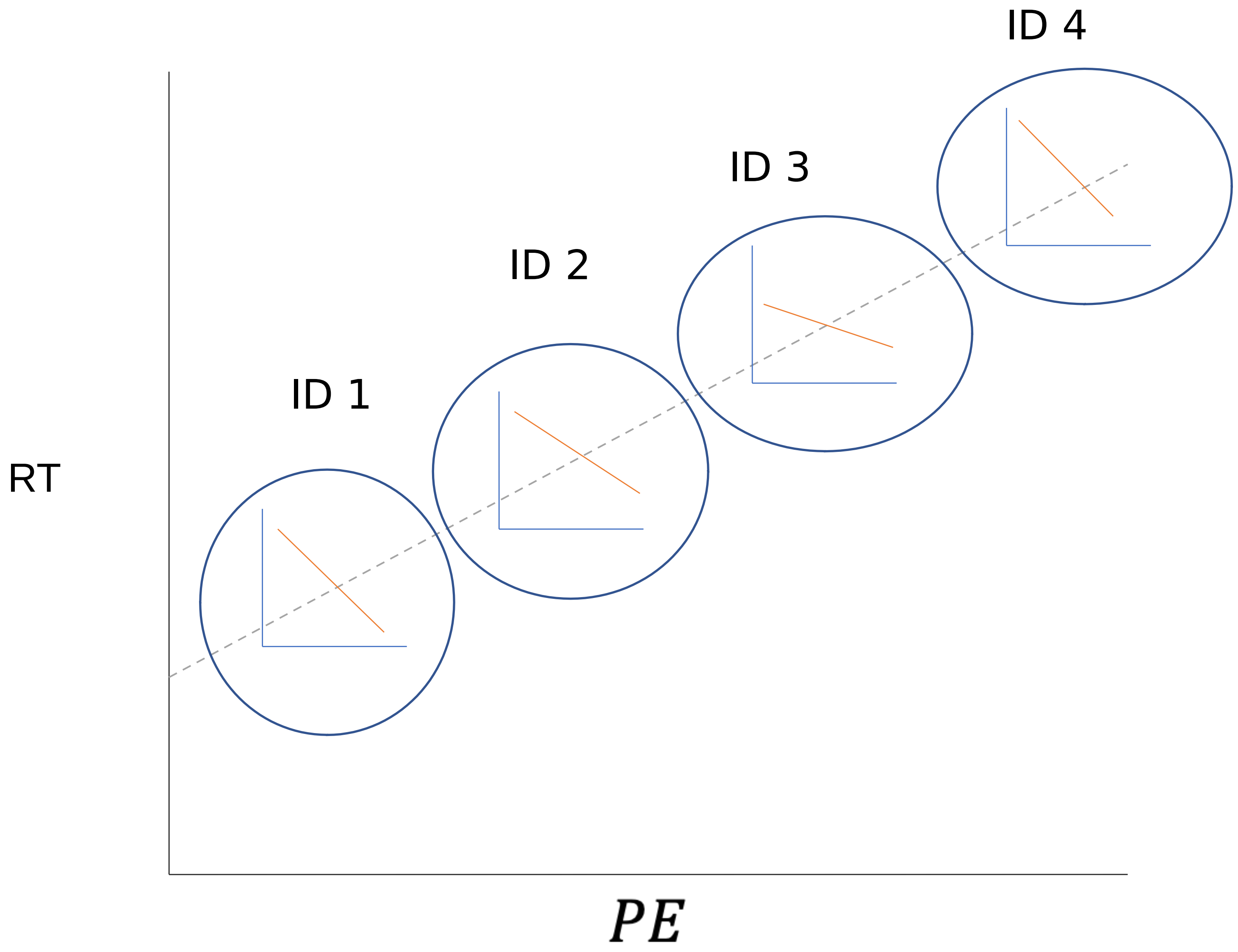


Linear mixed models
Multilevel models
Hierarchical linear models

Nested data

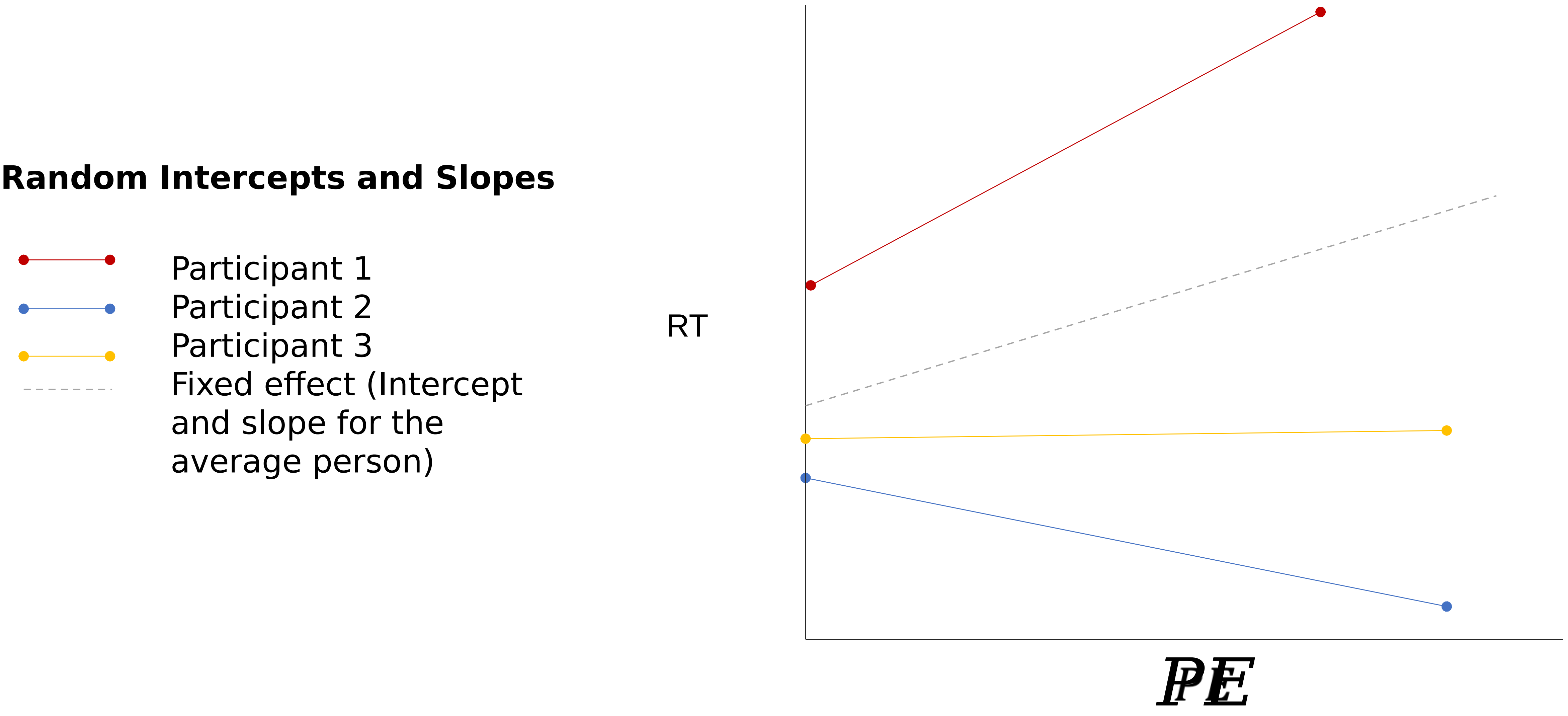
Taking into account within- and between-person variability





“The reason for favoring the within-subject level is that omitted and confounding variables are less likely to be a problem when analyses focus on how and why people change over time than on how people differ from one another”. (Bolger & Laurenceau, *Intensive longitudinal methods: An introduction to diary and experience sampling research, 2013*).

Random Intercepts and Slopes



$$Y_{ij} = \beta_{oj} + \beta_{1j} X_{ij} + \epsilon_{ij}$$

Level 1 - dependent variable Y for observation *i* for a particular subject *j.*

The dependent variable *Y* for subject *j* for the specific observation *i* is given by an subject-specific intercept β0j , a subject specific slope β1*j,* and a within-subject erro term εij.

$$Y_{ij} = \beta_{oj} + \beta_{1j} X_{ij} + \epsilon_{ij}$$

$$\beta_{oj} = \gamma_{00} + u_{0j}$$

Level 2 (between-subjects)

u0j is a random effect because it varies from cluster to cluster

Subject 1

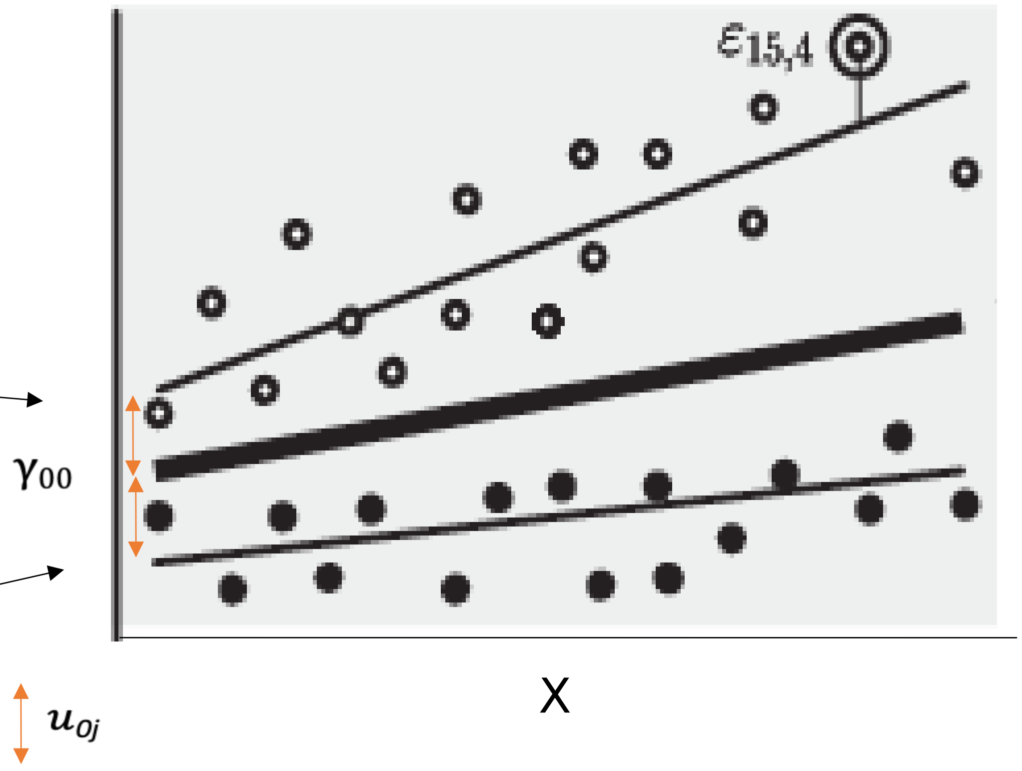
$$Y_{i1} = (\gamma_{00} + u_{01})$$

Typical person

$$Y_{ij} = \gamma_{00}$$

Subject 2

$$Y_{i2} = (\gamma_{00} + u_{02})$$



$$Y_{ij} = \beta_{oj} + \beta_{1j} X_{ij} + \epsilon_{ij}$$

$$\beta_{oj} = \gamma_{00} + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + u_{1j}$$

Level 2 (between-subjects)

Subject 1

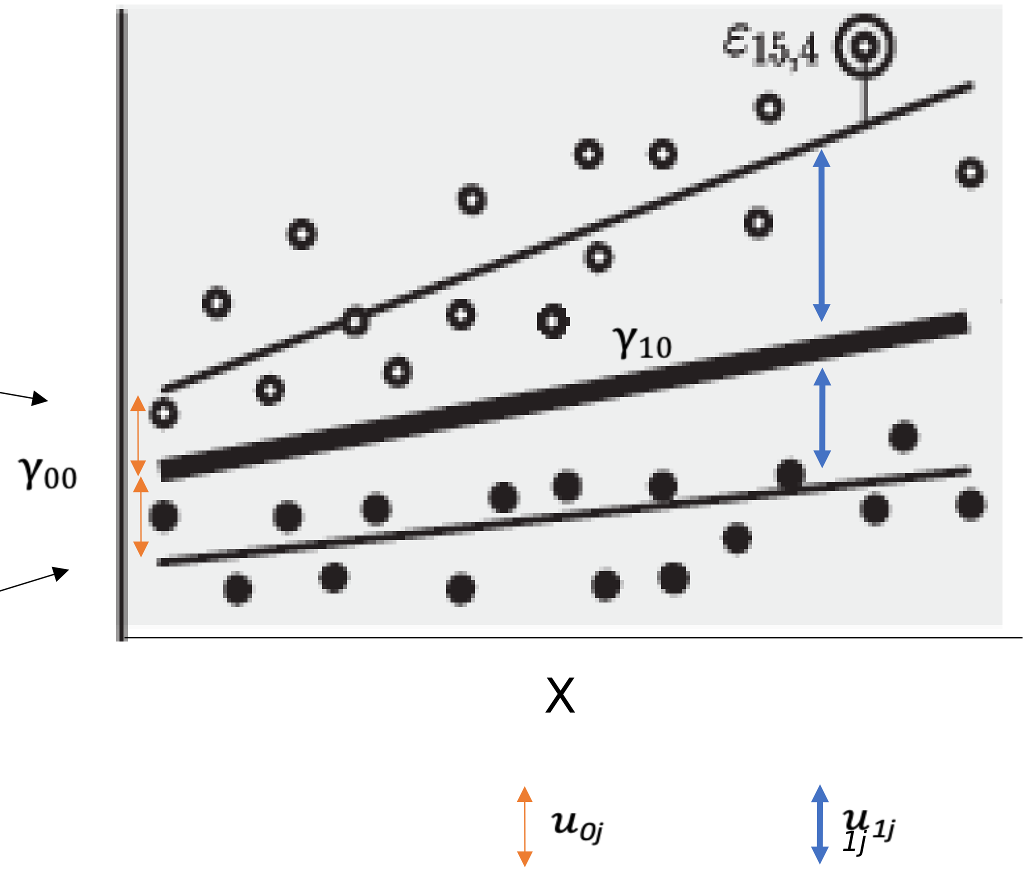
$$Y_{i1} = (\gamma_{00} + u_{01}) + (\gamma_{10} + u_{11})\, X_{i1} + \epsilon_{i1}$$

Typical person

$$Y_{ij} = \gamma_{00} + \gamma_{10} X_{ij}$$

Subject 2

$$Y_{i2} = (\gamma_{00} + u_{02}) + (\gamma_{10} + u_{12})\, X_{i2} + \epsilon_{i2}$$



$$Y_{ij} = \beta_{oj} + \beta_{1j} X_{ij} + \epsilon_{ij}$$

is a within-sbject residual term representing the difference, at a given time point between the predicted Y for a given subject and the actual value.

$$\beta_{oj} = \gamma_{00} + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + u_{1j}$$

Level 2 (between-subjects)

Subject 1

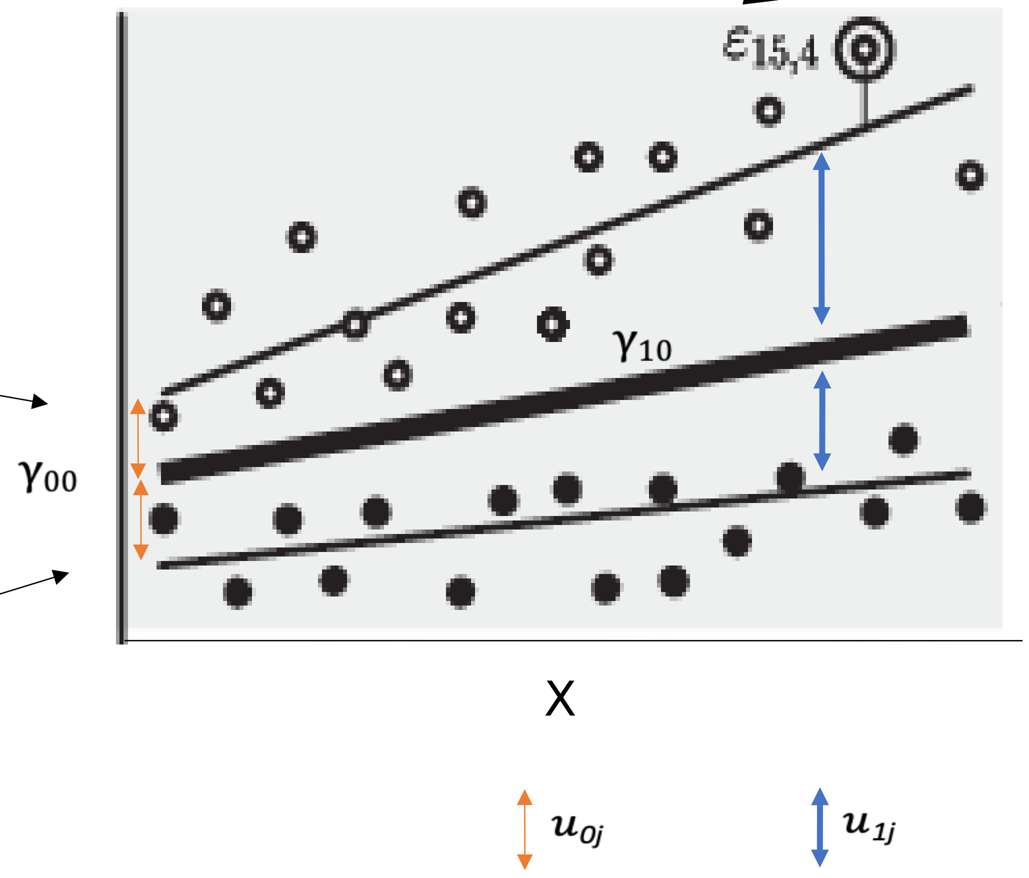
$$Y_{i1} = (\gamma_{00} + u_{01}) + (\gamma_{10} + u_{11})\, X_{i1} + \epsilon_{i1}$$

Typical person

$$Y_{ij} = \gamma_{00} + \gamma_{10} X_{ij}$$

Subject 2

$$Y_{i2} = (\gamma_{00} + u_{02}) + (\gamma_{10} + u_{12})\, X_{i2} + \epsilon_{i2}$$



How are parameters estimated?

Parameters for mixed models are not estimated through OLS, which is not the optimal approach for complex model. Instead, they are estimated through Maximum Likelihood Estimation (MLE) and Restricted Maximum Likelihood (REML).

MLE search for the population model parameters that maximize the likelihood of obtaining our data. In other words, the parameters obtained should maximize the likelihood of our particular sample.
It search through candidate parameters in several iterations using nonlinear optimization algorhitms.

RMLE is like MLE, but also takes into account the number of parameters being estimated in the model in order to determine the appropriate degrees of freedom for the estimation of the random components. In contrast, MLE does not account for these. Therefore, it is generally preferred for estimating multilevel models. It is the default for lmer.

Let’s do this!

```
lmm<-lmer(DV~IV+(randomeffect|randomintercept),data=df)
```

Random, “stochastic” part of the model, with the random effects. On the right of the | symbol there is the clustering variable. On te left side there are the random slopes.

First, let’s fit an unconditional model, a model with only the intercept, to examine the variability between- and within- participant

$$\rho_I = \frac{\tau^2}{\tau^2 + \sigma^2}$$

where
τ^2 = Population variance between clusters
σ^2 = Population variance within clusters

Relative higher values of ρ indicate that great amount of variation in the outcome measure (DV) is associated with the cluster membership – observation within participants are correlated.
Variance between-participant is higher that within-participant.

This is called intraclass correlation.

Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']
Formula: RT ~ 1 + (1 | subj_id)

Lmmworkshop.Rmd

```{r lmm rand int}
```

```
Random effects:
 Groups   Name        Variance Std.Dev.
 subj_id  (Intercept)  3.0549   1.7478
 Residual              0.2816   0.5307
Number of obs: 9000, groups:  subj_id, 30
```

$$\hat{\rho} = \frac{3.05}{3.05 + 0.28} = 0.91$$

To test whether the addition of the random intercept improves the fit of the model, we can compare a model with random intercept and a model without them

````
Lmmworkshop.Rmd
```{r lmm rand test sig}
```
````

This is called “likelihood ratio test”, where the fit of the full and reduced models are compared.

The glmer effect should be called first. anova(mixmod_unc, mod_unc)

```
Data: df
Models:
mod_unc: RT ~ 1
mixmod_unc: RT ~ 1 + (1 | subj_id)
           npar   AIC   BIC   logLik deviance Chisq Df Pr(>Chisq)
mod_unc       2 36110 36124 -18053.1    36106
mixmod_unc    3 14385 14406  -7189.3    14379 21728  1  < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

If models are not fitted with ML, they are refitted with ML, because it makes them comparable.

The chi square difference test is computed on deviance statistics, which is -2 *LL, and follows a chi-square distribution.

Not everyone agrees that testing significance of random effects is a good idea, as the LRT in some cases are conservative
Checkout this:

https://bbolker.github.io/mixedmodels-misc/glmmFAQ.html#testing-significance-of-random-effects

Consider not testing the significance of random effects. If the random effect is part of the experimental design, this procedure may be considered 'sacrificial pseudoreplication' (Hurlbert 1984). Using stepwise approaches to eliminate non-significant terms in order to squeeze more significance out of the remaining terms is dangerous in any case.

consider using the RLRsim package, which has a fast implementation of simulation-based tests of null hypotheses about zero variances, for simple tests. (However, it only applies to lmer models, and is a bit tricky to use for more complex models.)

Lmmworkshop.Rmd

```{r exactLRT}
```

```
LM<-lm(Acc~PE,data=data)
```

Function that stands for “linear model” and is used in R to calculate regression and its special cases (Anova, multiple regressions, etc.)

Outcome, or dependent variable

On the right side of the ~ symbol there are the predictors (fixed effects)

```
Library("lme4")
MLM<-lmer(Acc~PE+(PE|PartNumb), data=data)
```

Function for the linear mixed model. It requires that some random effects are addded.

Fixed effects part. Same as the lm one.

Within brackets there is the “stochastic” part of the model, with the random effects.
On the right of the | symbol there is the clustering variable: in our case is the participant number variable. We are adding random intercepts for participants, meaning that we are considering their variance in the average recognition accuracy. On the left side of the | symbol there are the random slopes. In this case, we are considering: random slopes for PE; this means that we are accounting for participants differences in the effects of this variable on accuracy

```
summary(MLM)

Random effects:
 Groups   Name        Variance  Std.Dev. Corr
 PartNumb (Intercept) 0.3662782 0.60521
          PE          0.0008159 0.02856  -1.00

 Residual             1.2540770 1.11986
Number of obs: 627, groups:  PartNumb, 38

Fixed effects:
            Estimate Std. Error        df t value Pr(>|t|)
(Intercept) -0.232759   0.133371 41.963984  -1.745
0.0883 .
PE    1       0.011369   0.006011 40.291277   1.891
0.0658 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Between participant’s variance (and std. dev)in the intercept (overall accuracy)

Between participant’s variance (and std. Dev) in the slope (effect of PE on accuracy).

Fixed effects. Intercept and coefficient for the entire sample, after accounting for between participants variance.

To sum up, benefits of using LMM:

- Control for random variability between items
- Focus on within-participants processes
- Separate within-participant from between-participant variability
- Account for dependency in the data
- Allow to deal with missing data

Please check the practical example in the Lab Tutorials for more on LMM