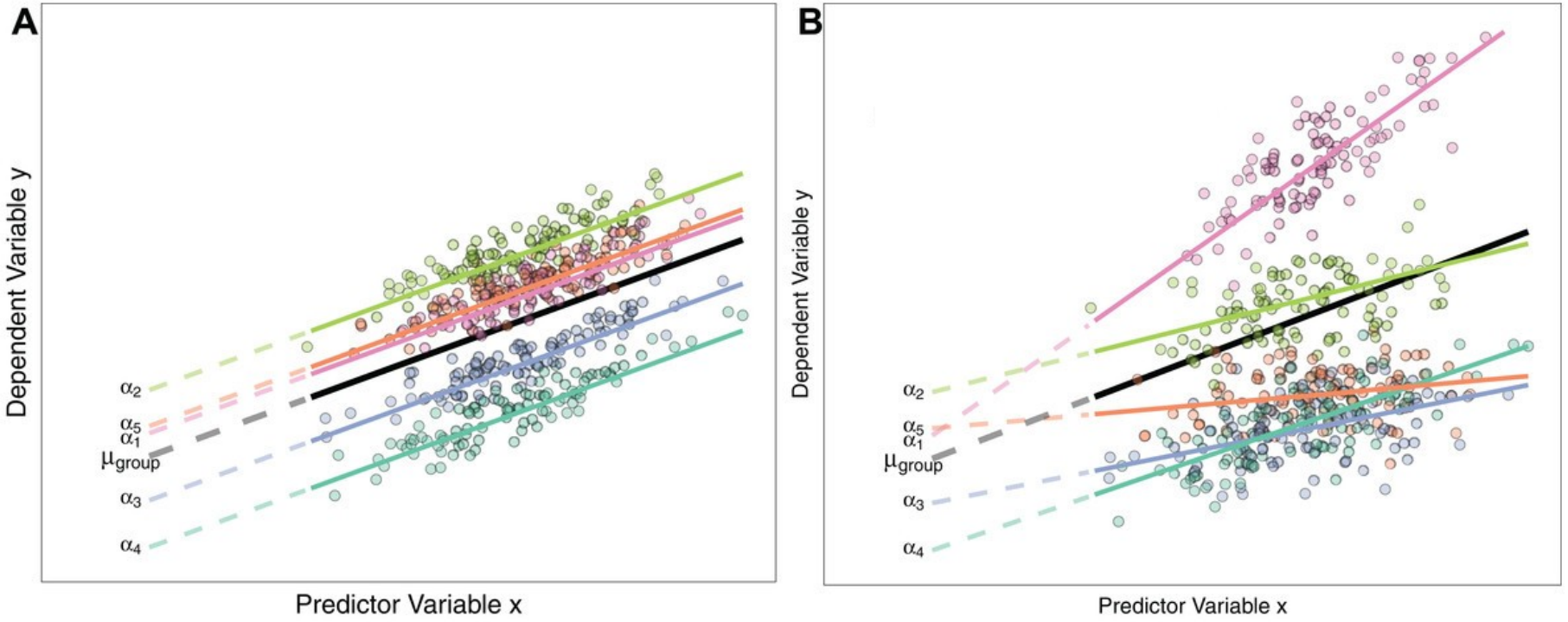
Introduction to:

Linear Mixed Effects models



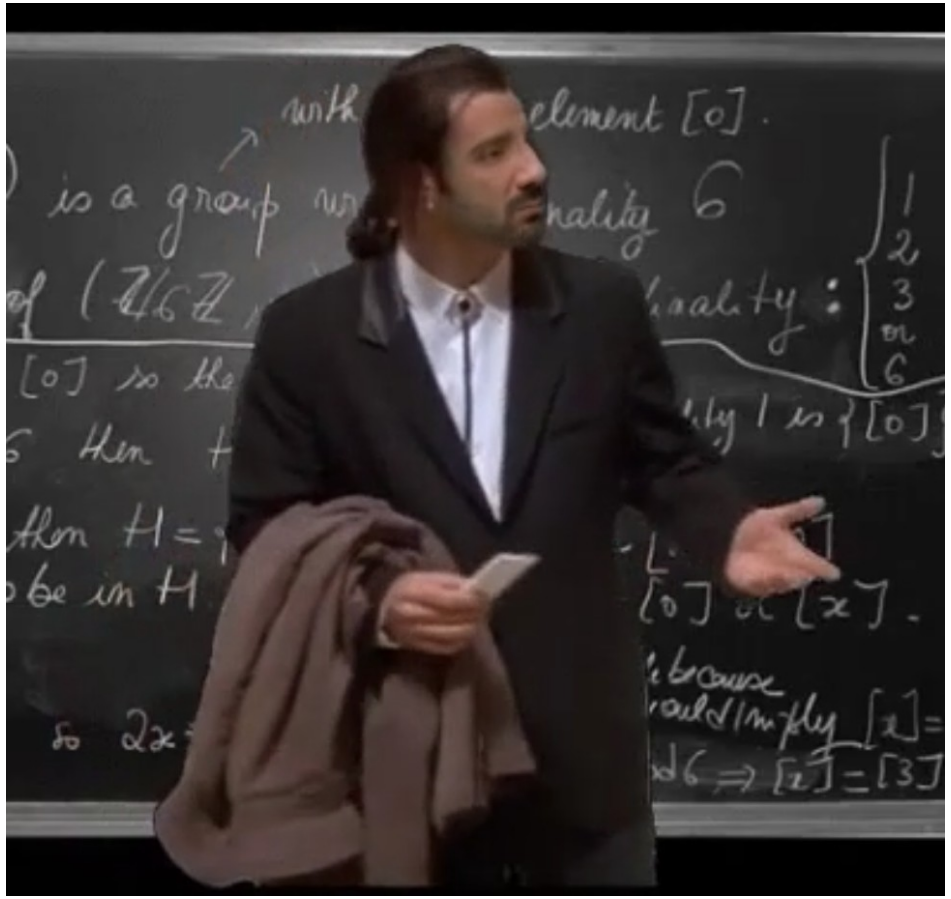
Linear mixed models
Linear mixed-effects models
Multilevel models
Multilevel regression
Multilevel linear modeling
Hierarchical linear models
Random coefficient models

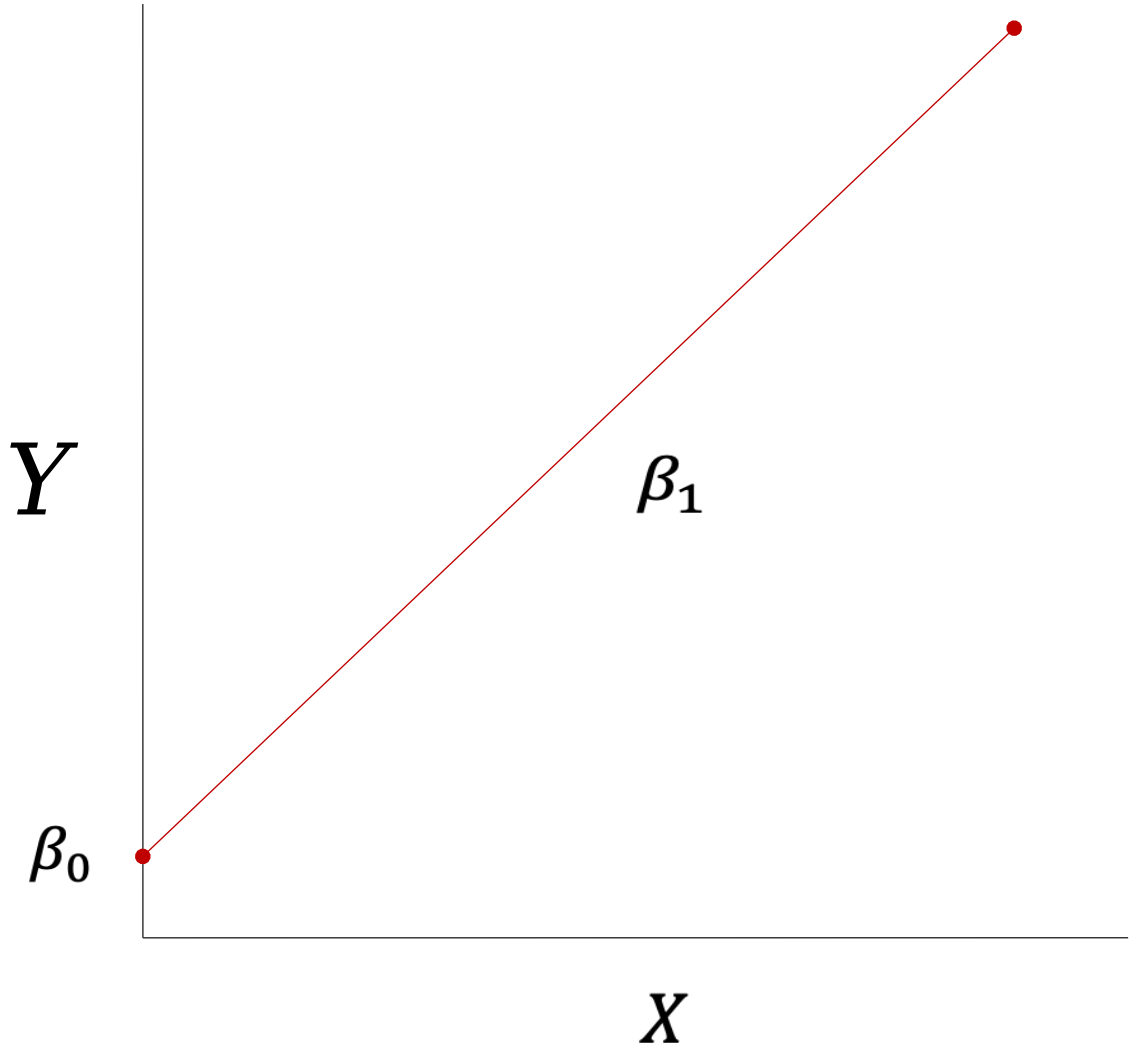
What are we goint to cover?

- Linear models: the basics
- Linear mixed - effects models: random effects
- Parameter estimation
- Intraclass correlation
- Calculating R^2 within- between
- Variance - covariance matrices
- Model selection
- Linear models with categorical predictors
- Setting contrasts
- Linear models with binary outcome (generalised linear models)
- Troubleshooting
- Reporting results

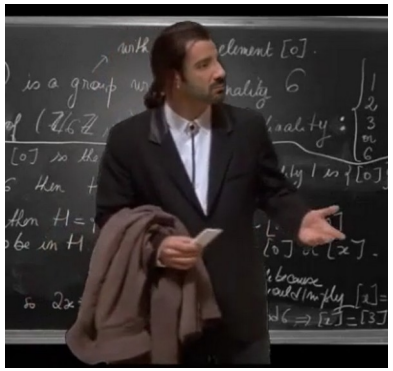
Theory Practice Excercises

Linear models

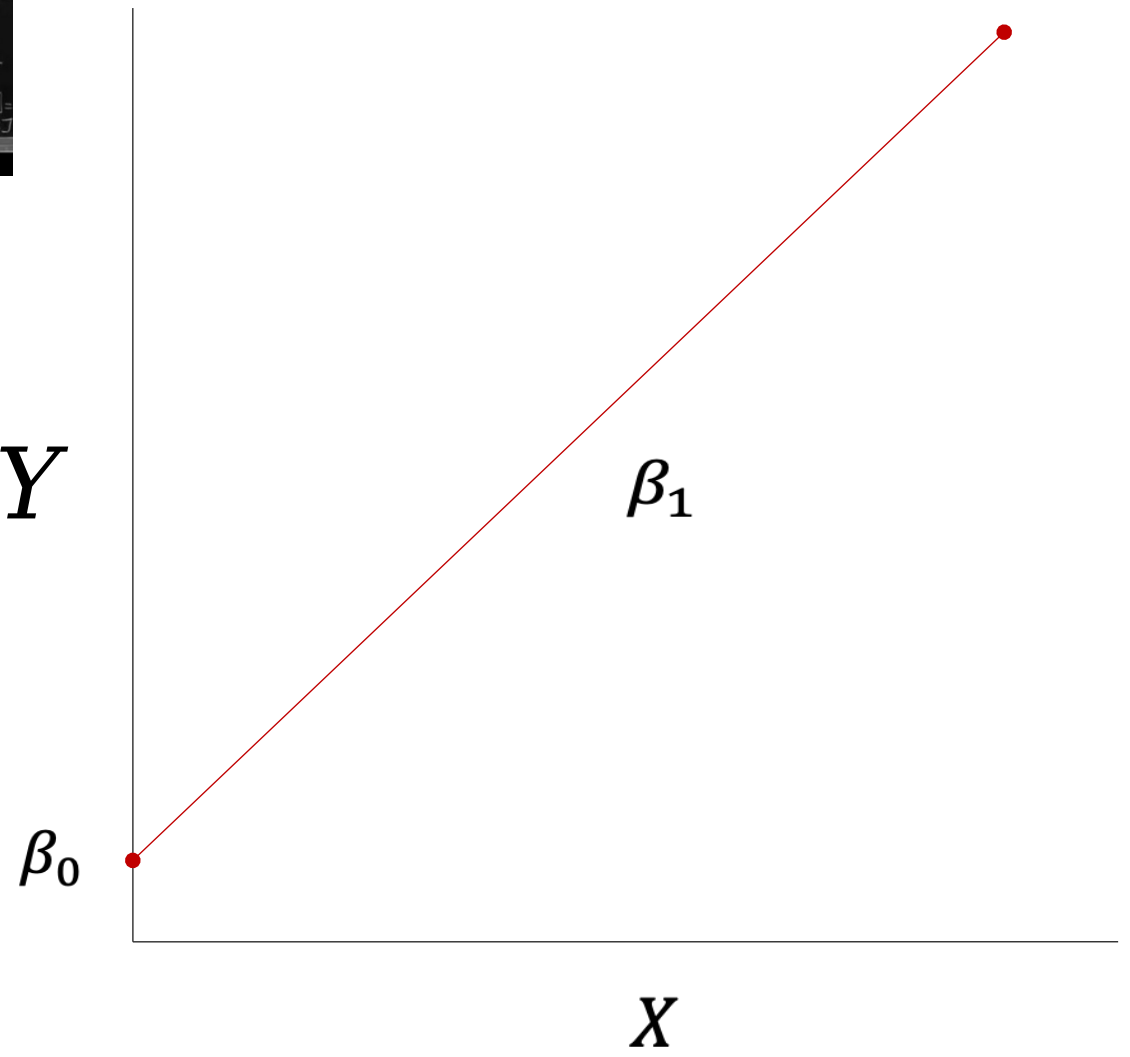




Linear models

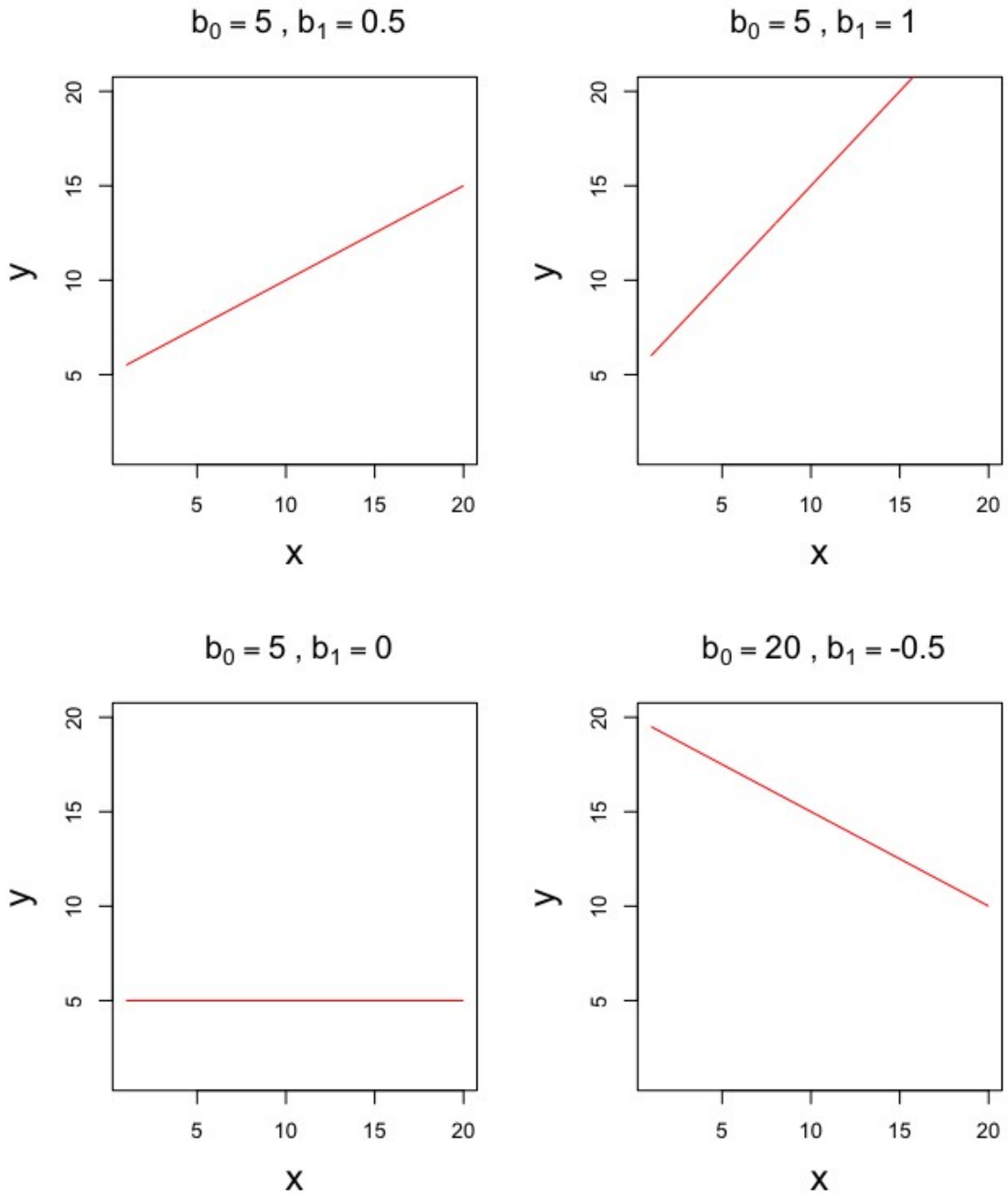


$$Y = \beta_0 + \beta_1 X$$



Linear models

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

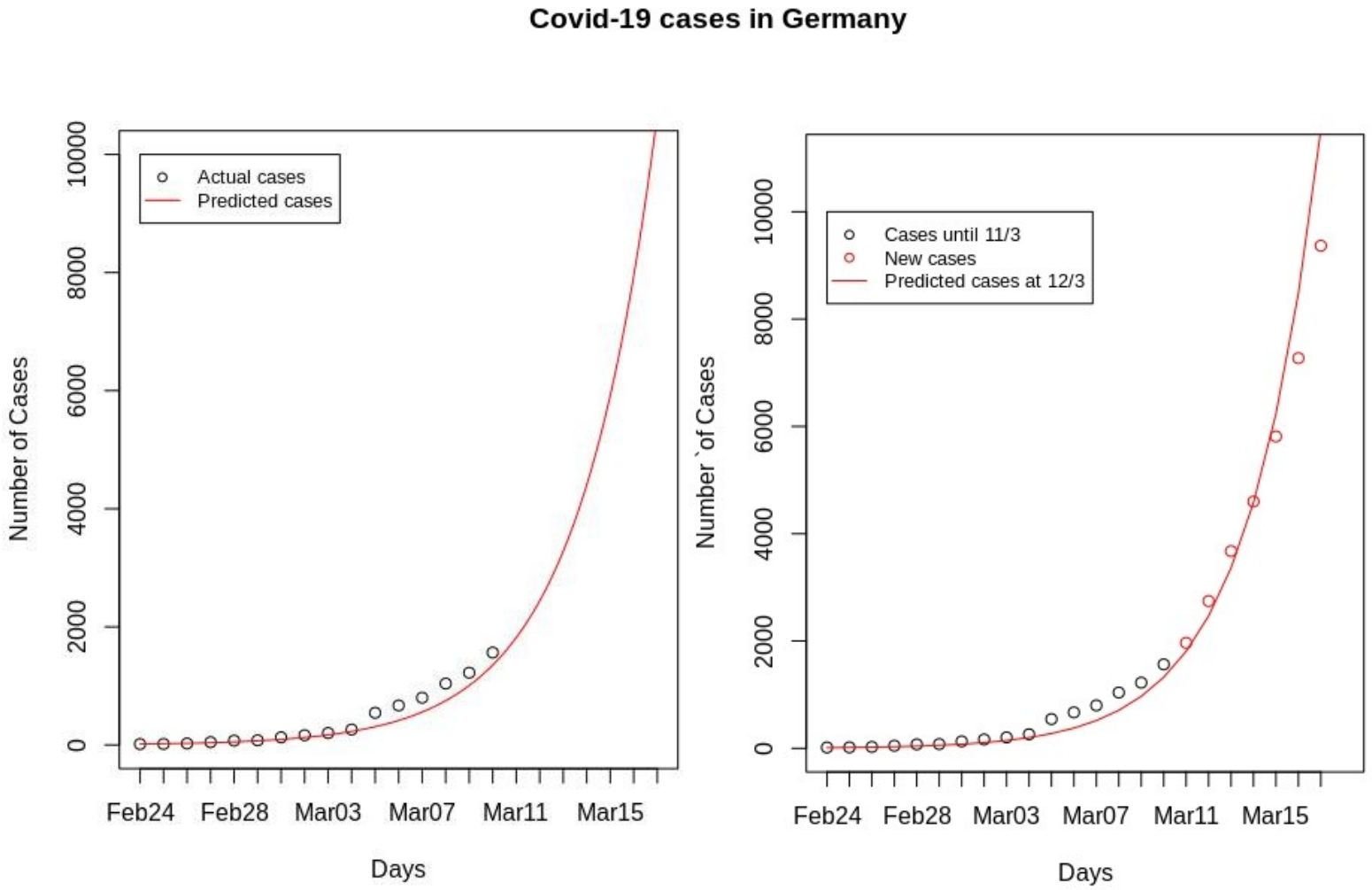


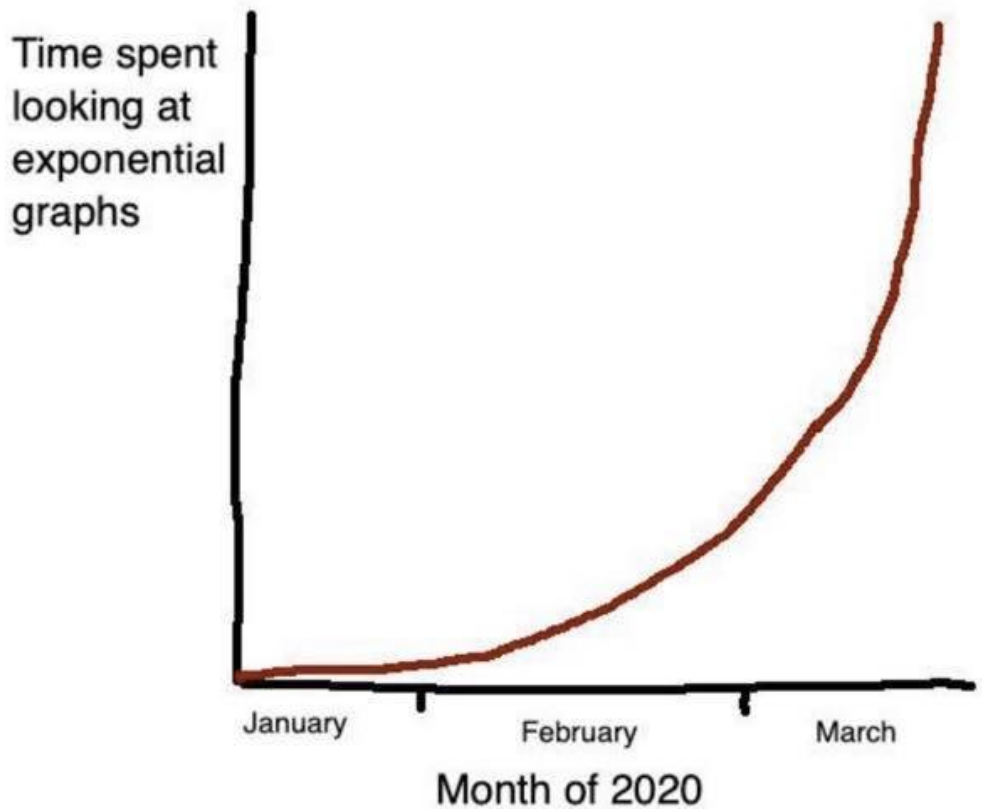
Linear models

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

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

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

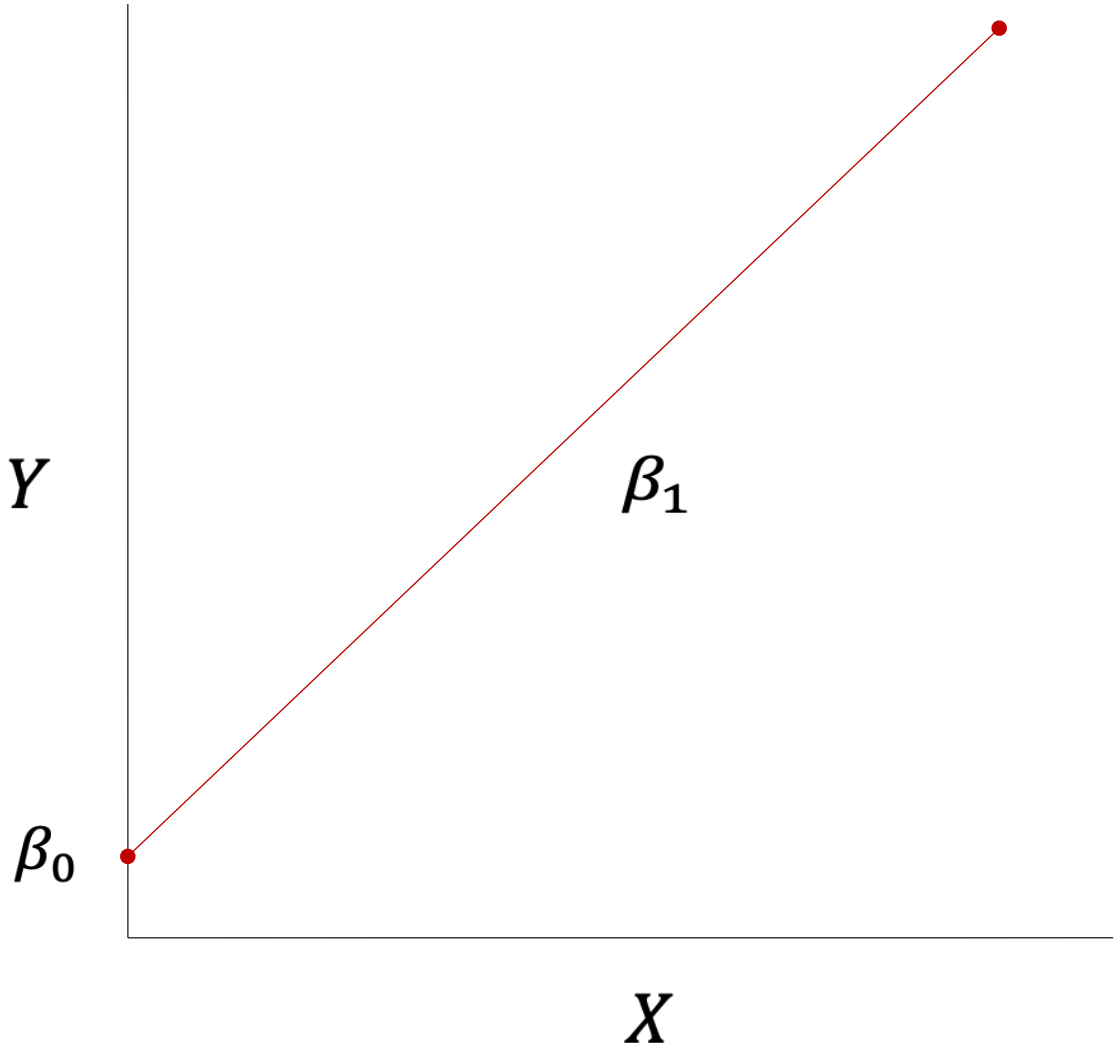



Time spent looking at exponential graphs
January
February
March
Month of 2020

Linear 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.

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

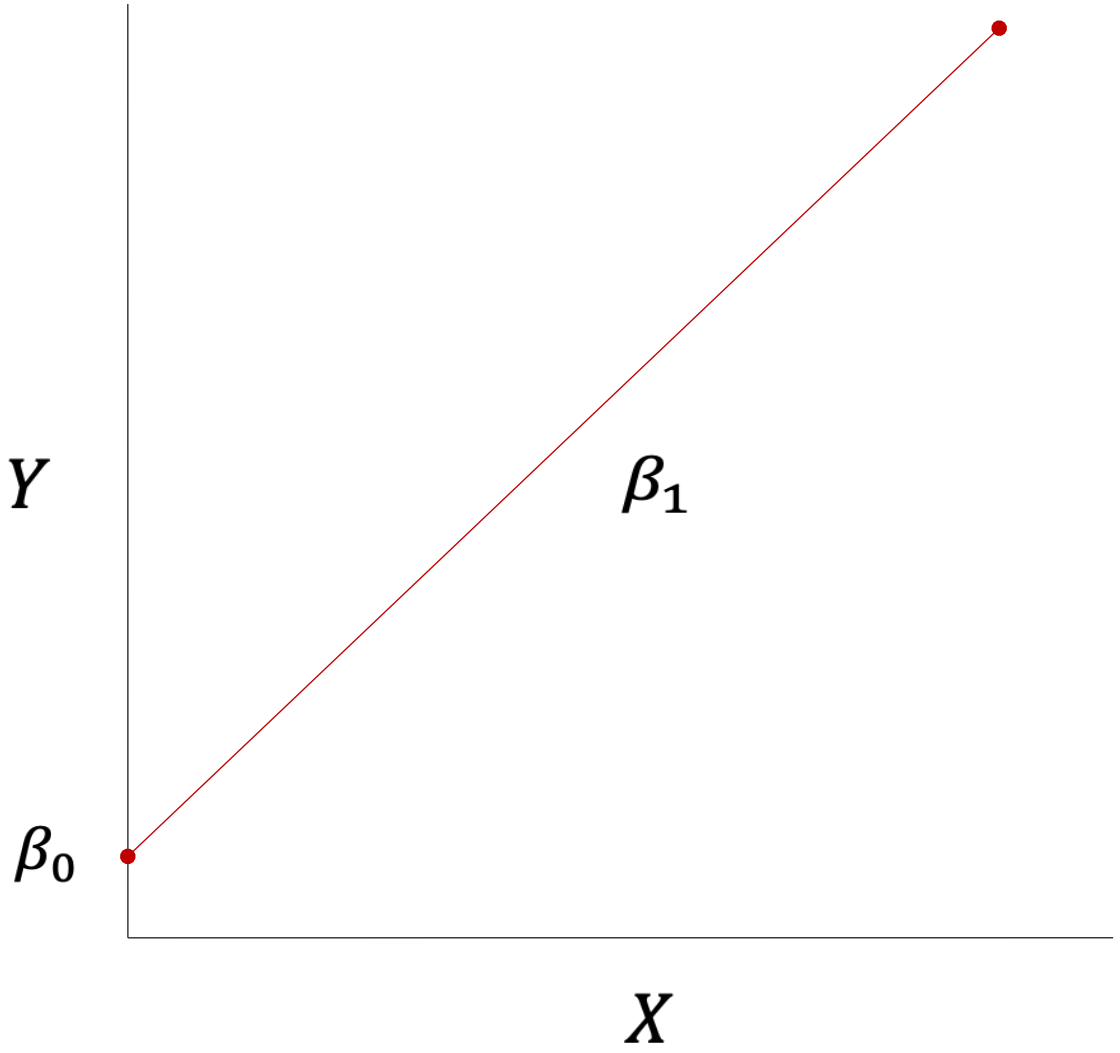


Linear models

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

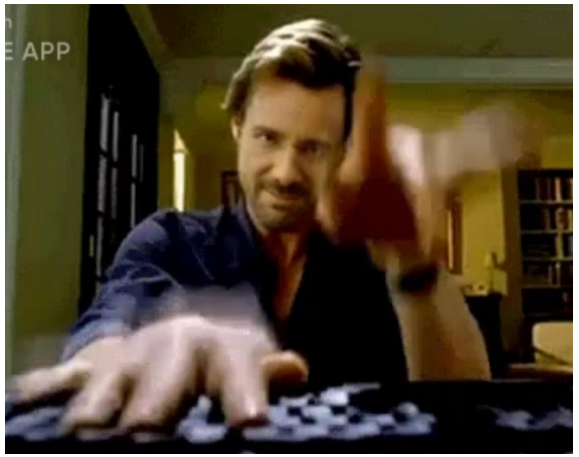
In linear models, the parameters β_0 and β_1 are estimated through OLS.

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.



Linear models

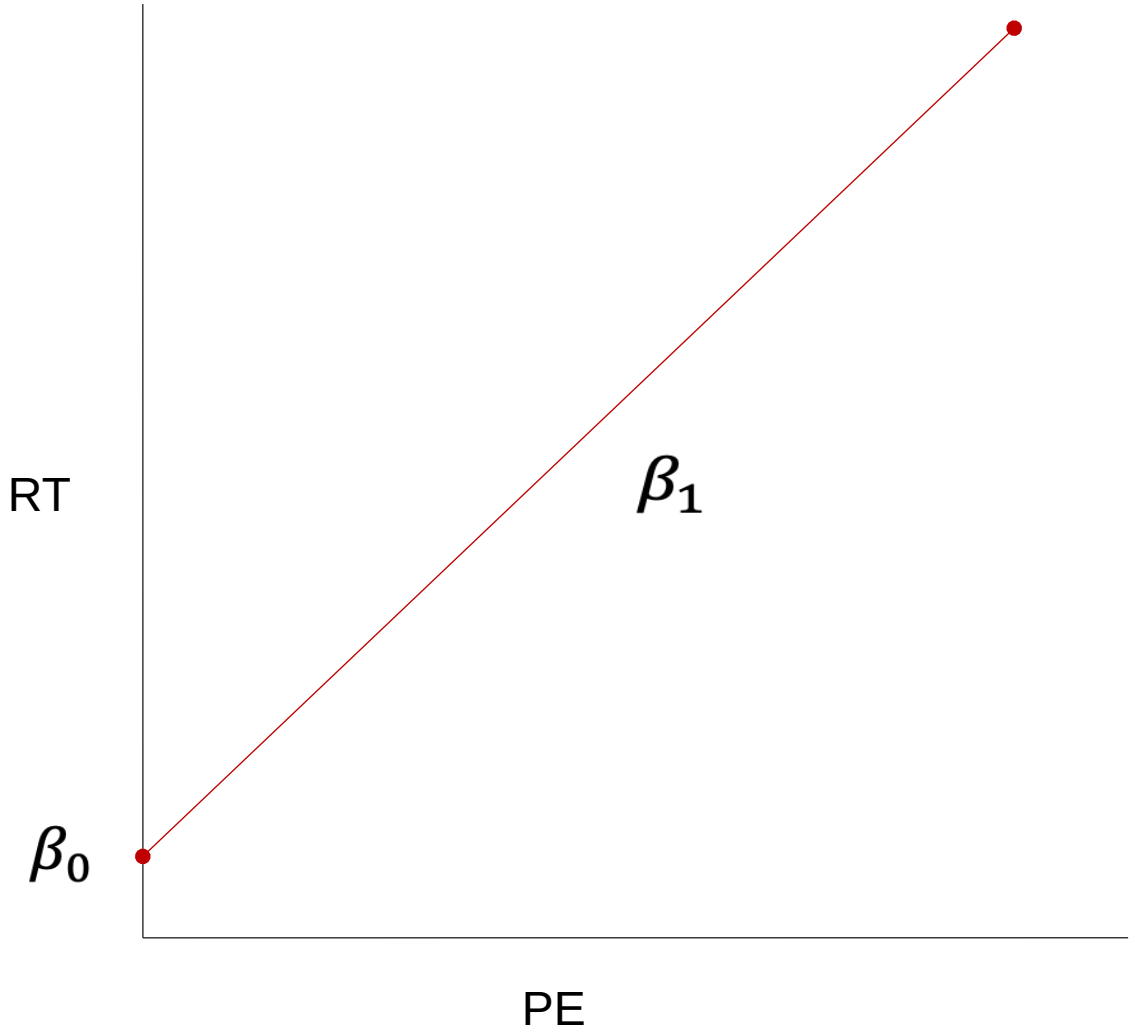
What if we have variables that vary within-participants?



Lmmworkshop.Rmd

```{r simulate Data}
```

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

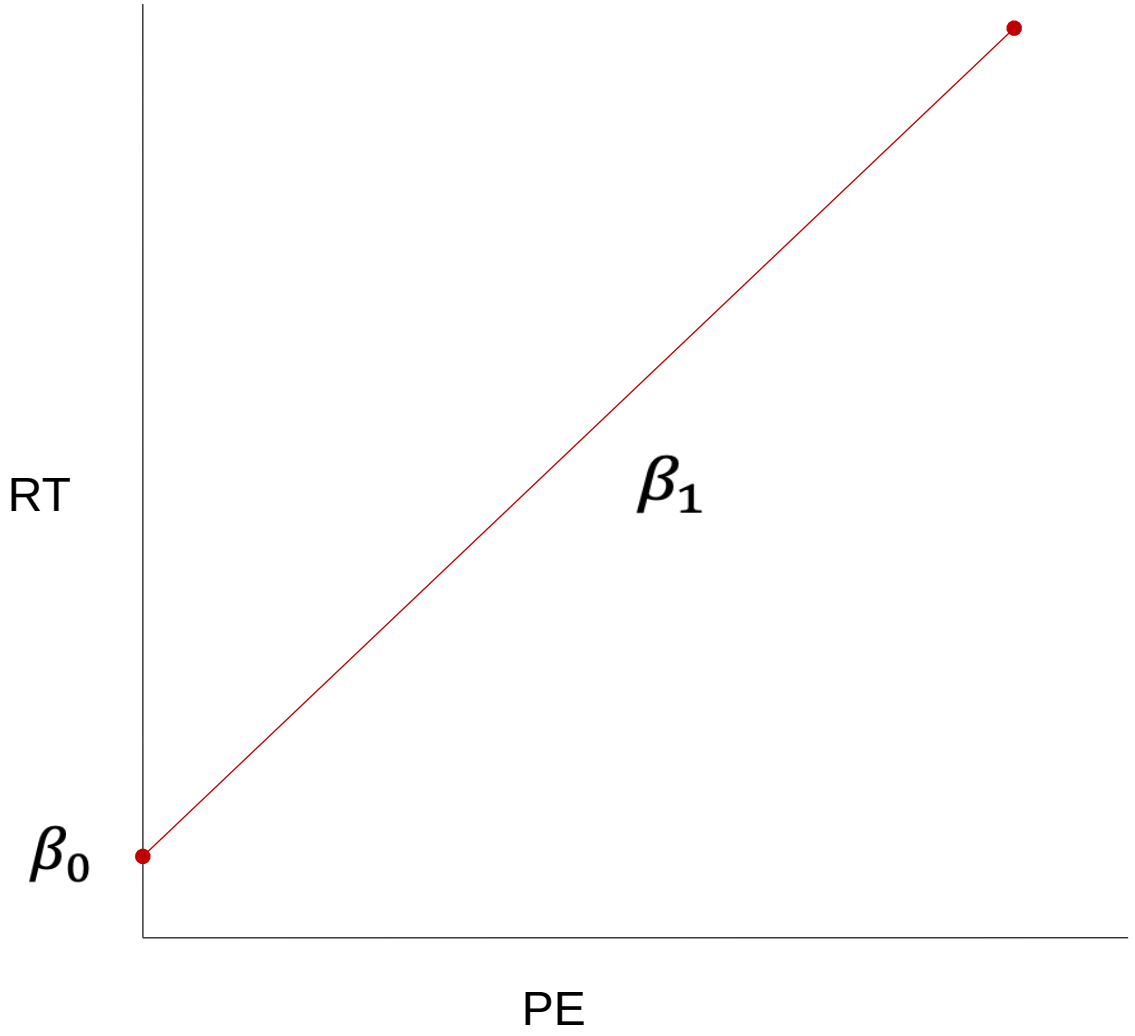


Linear models

What if we have variables that vary within-participants?

subj_id	item_id	PE	RT
1	1	0.25201203	6.038006
1	2	0.38320504	5.338470
1	3	0.58949096	5.296091
1	4	0.36118640	5.962406
1	5	0.59537556	4.633638
1	6	0.48636174	5.450406
1	7	-0.21139284	7.370716
1	8	0.85326956	4.480162
1	9	0.46687076	5.201361
1	10	0.83353662	4.918313
1	11	-0.41113802	7.658535
1	12	-0.68435719	8.654607
1	13	0.03561061	6.484362
1	14	-0.34669939	7.299588
1	15	-0.49267744	7.641731
1	16	-0.58752750	8.057258
1	17	0.49731023	5.309548
1	18	0.67185395	5.253346

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



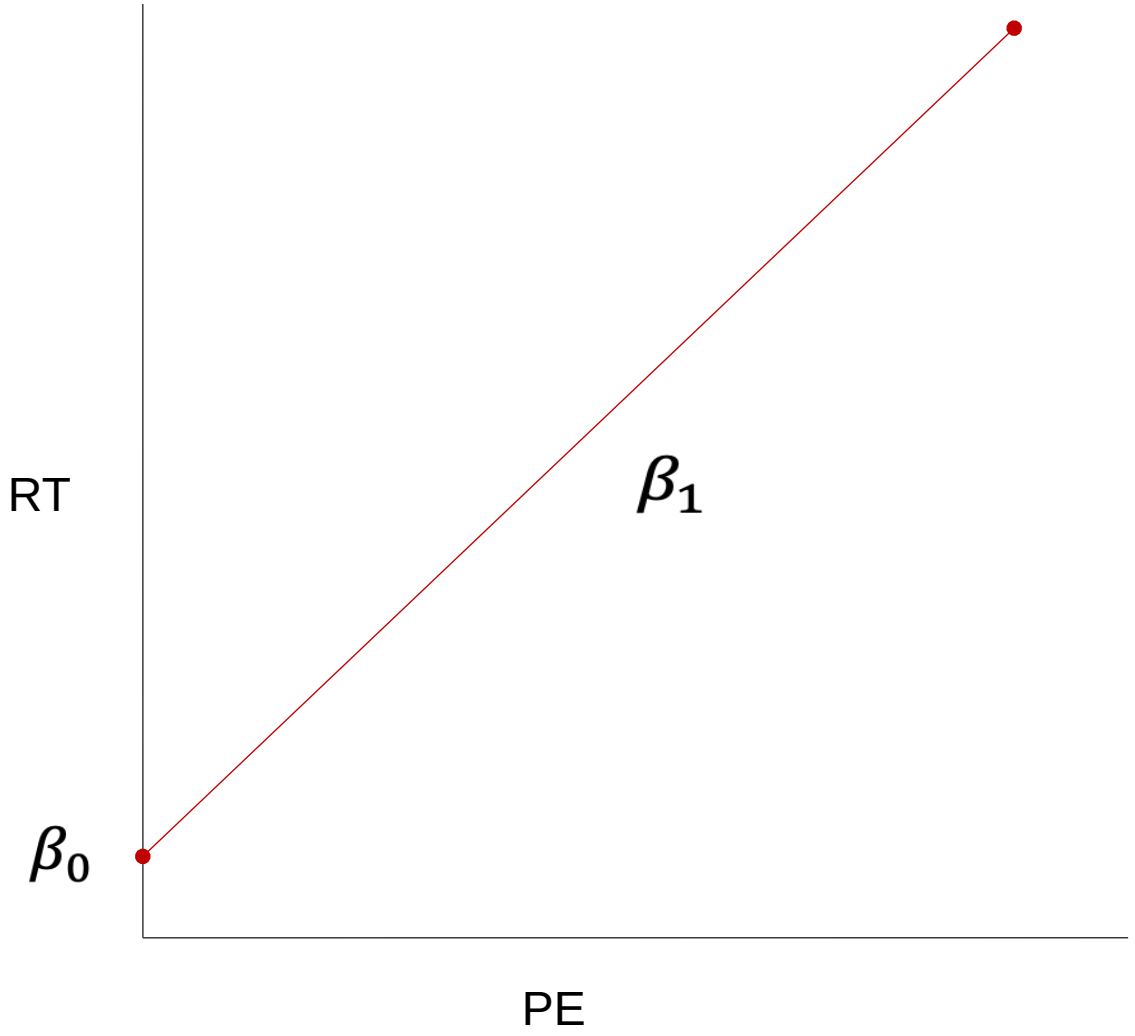
Linear models

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

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).



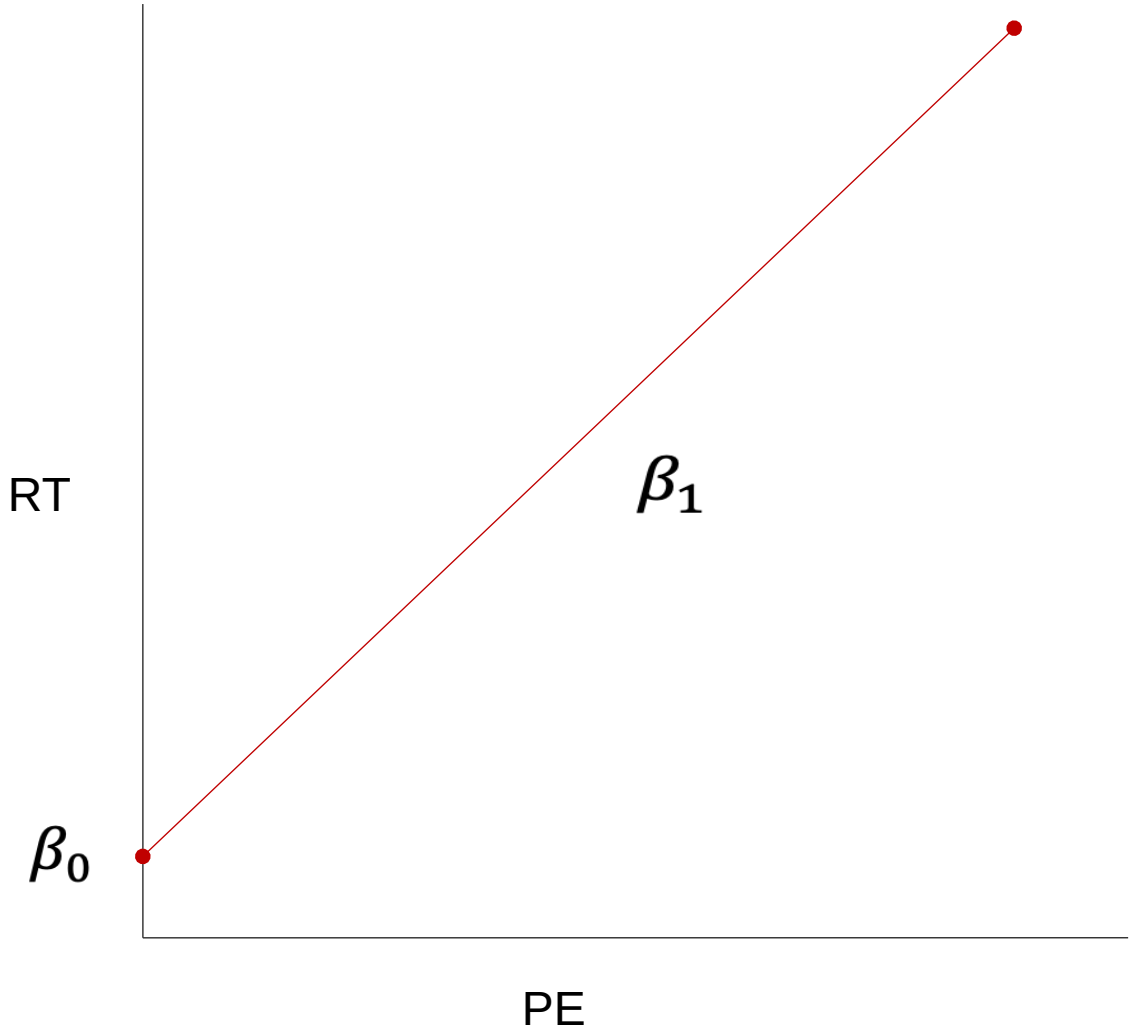
Linear models

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

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

Lmmworkshop.Rmd

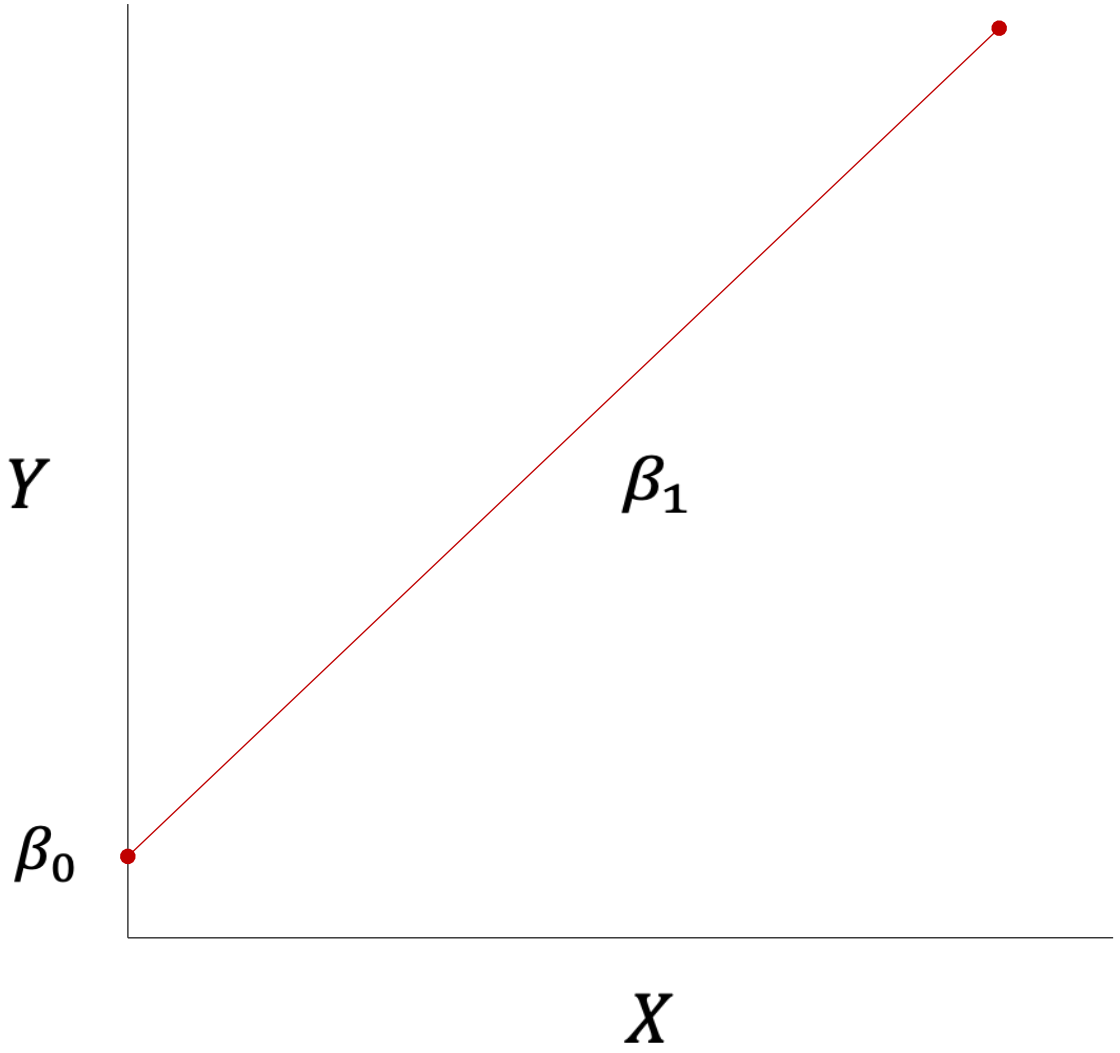
```{r aggregate }
```

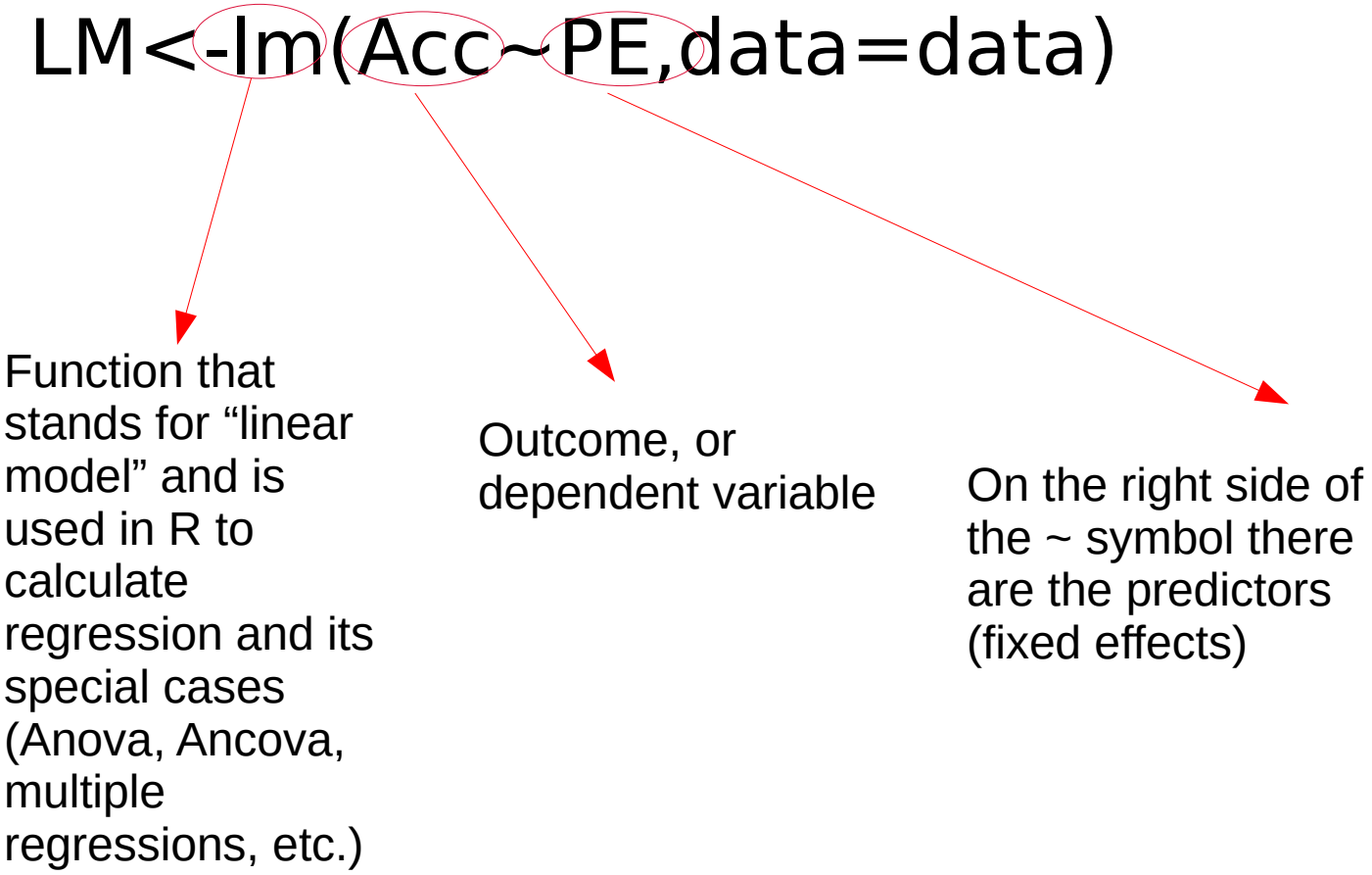


Linear models

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

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




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, Ancova, multiple regressions, etc.)
Outcome, or dependent variable
On the right side of the ~ symbol there are the predictors (fixed effects)

Linear models

summary(linearmodel)

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

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

Residuals:

Min	1Q	Median	3Q	Max
-2.7819	-1.5765	0.0868	0.9321	5.3083

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	8.233	7.024	1.172	0.251
PE	-3.730	48.573	-0.077	0.939

Residual standard error: 1.905 on 28 degrees of freedom
Multiple R-squared: 0.0002105, Adjusted R-squared: -0.0355
F-statistic: 0.005897 on 1 and 28 DF, p-value: 0.9393

Linear models

summary(linearmodel)

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

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

Residuals:

Min	1Q	Median	3Q	Max
-2.7819	-1.5765	0.0868	0.9321	5.3083

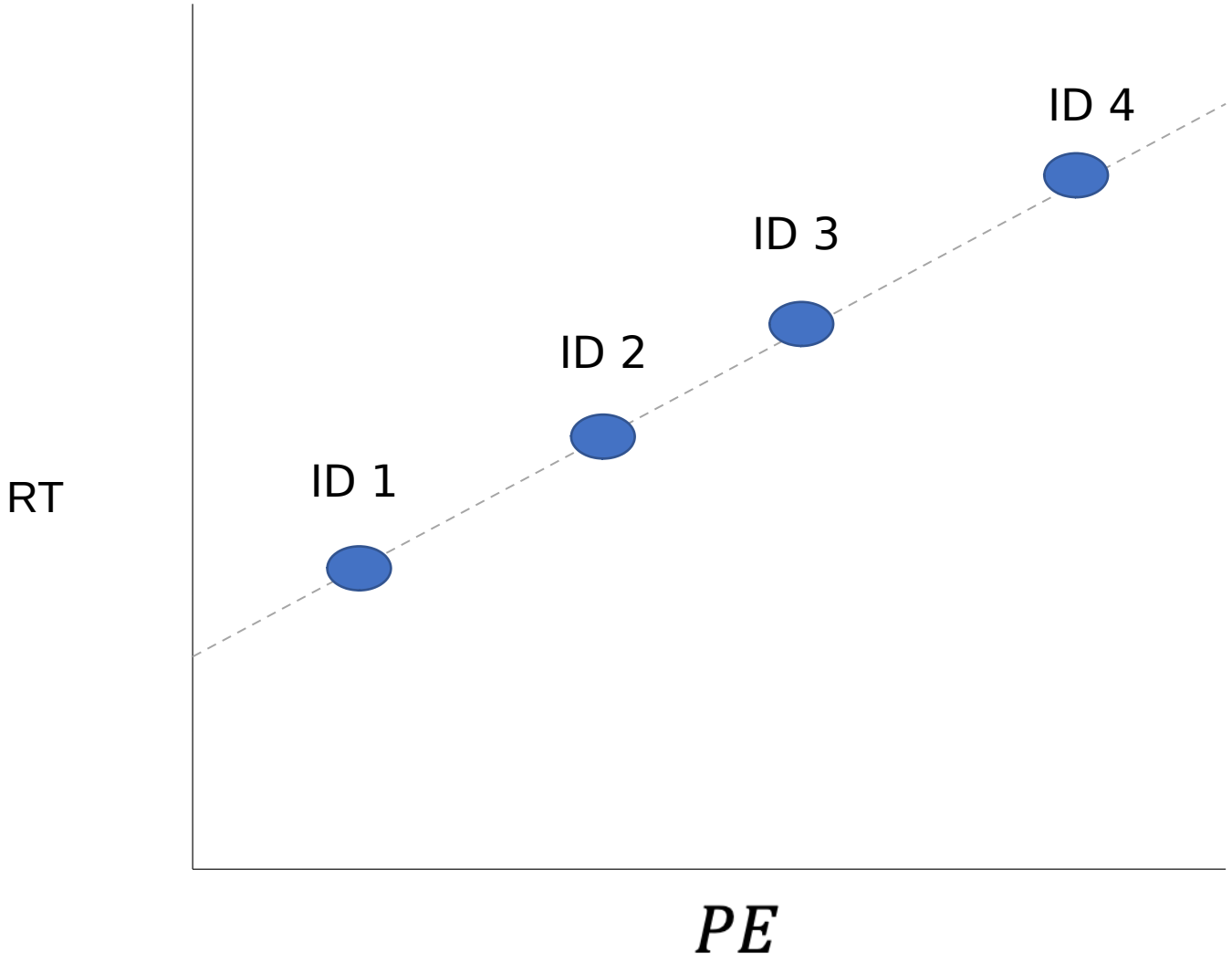
Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	8.233	7.024	1.172	0.251
PE	-3.730	48.573	-0.077	0.939

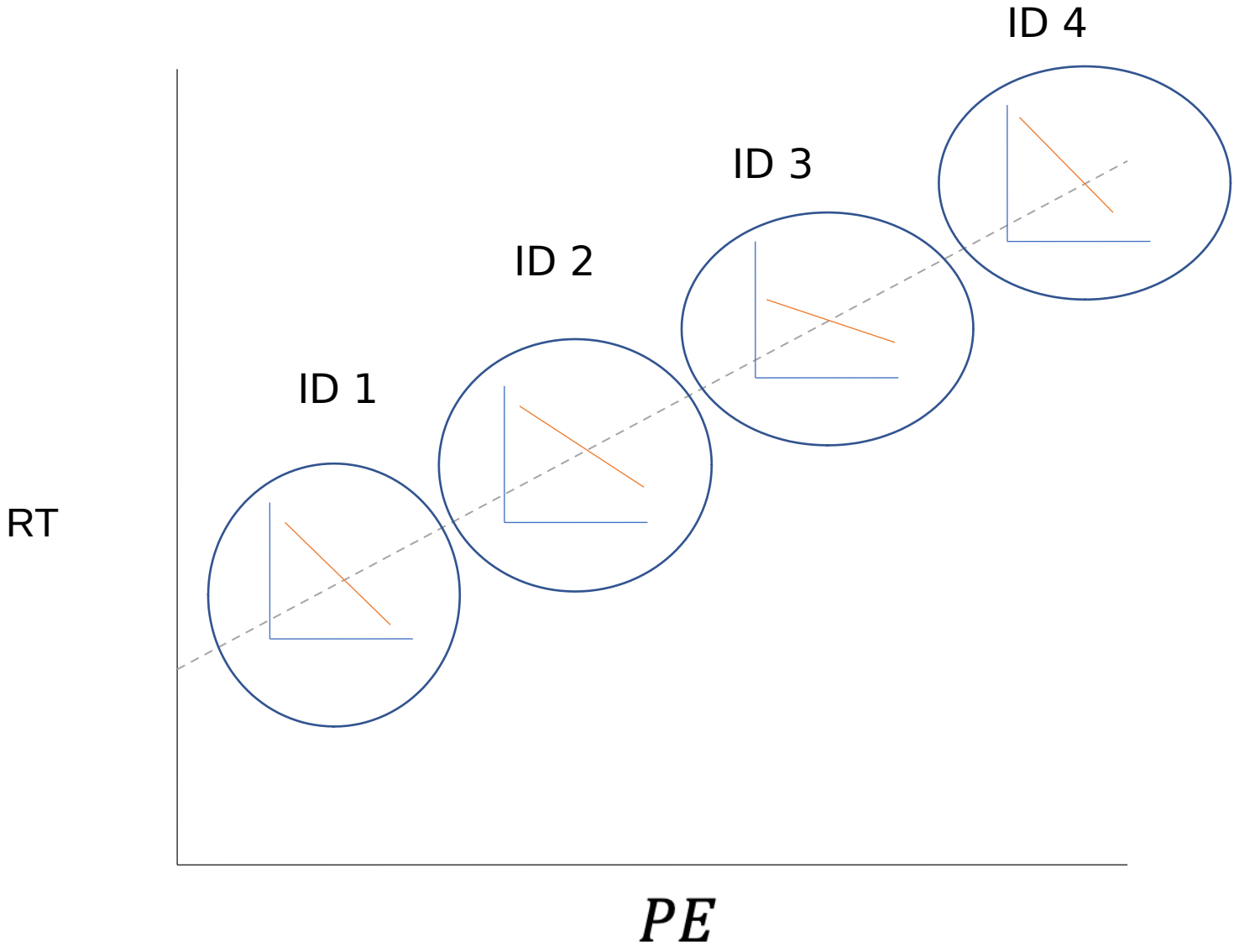
Residual standard error: 1.905 on 28 degrees of freedom
Multiple R-squared: 0.0002105, Adjusted R-squared: -0.0355
F-statistic: 0.005897 on 1 and 28 DF, p-value: 0.9393

Linear models

What is the problem with that?



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

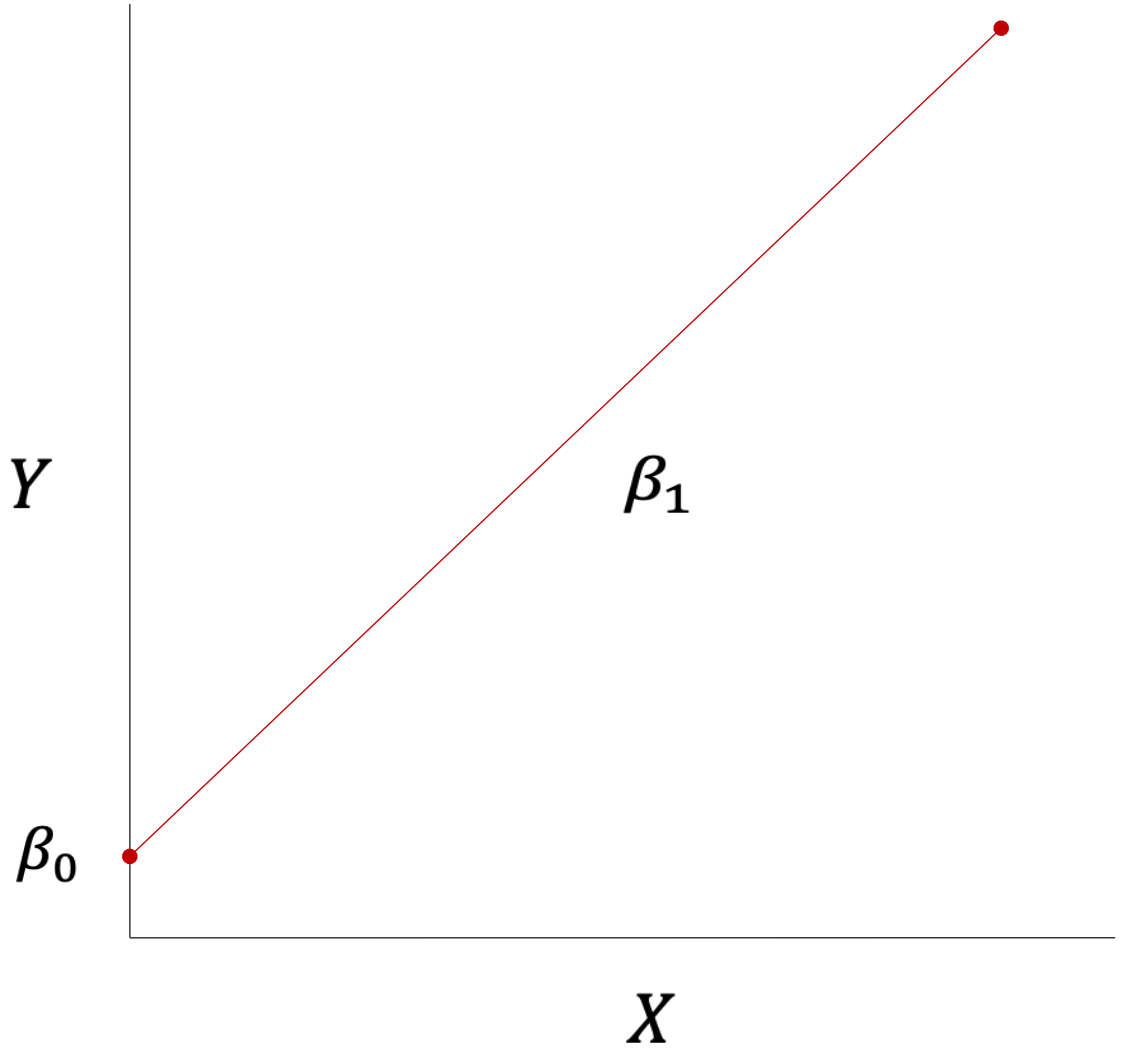


Within-participant level:
When participants experienced higher PE their RT are faster

Linear models

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

So what do we do?

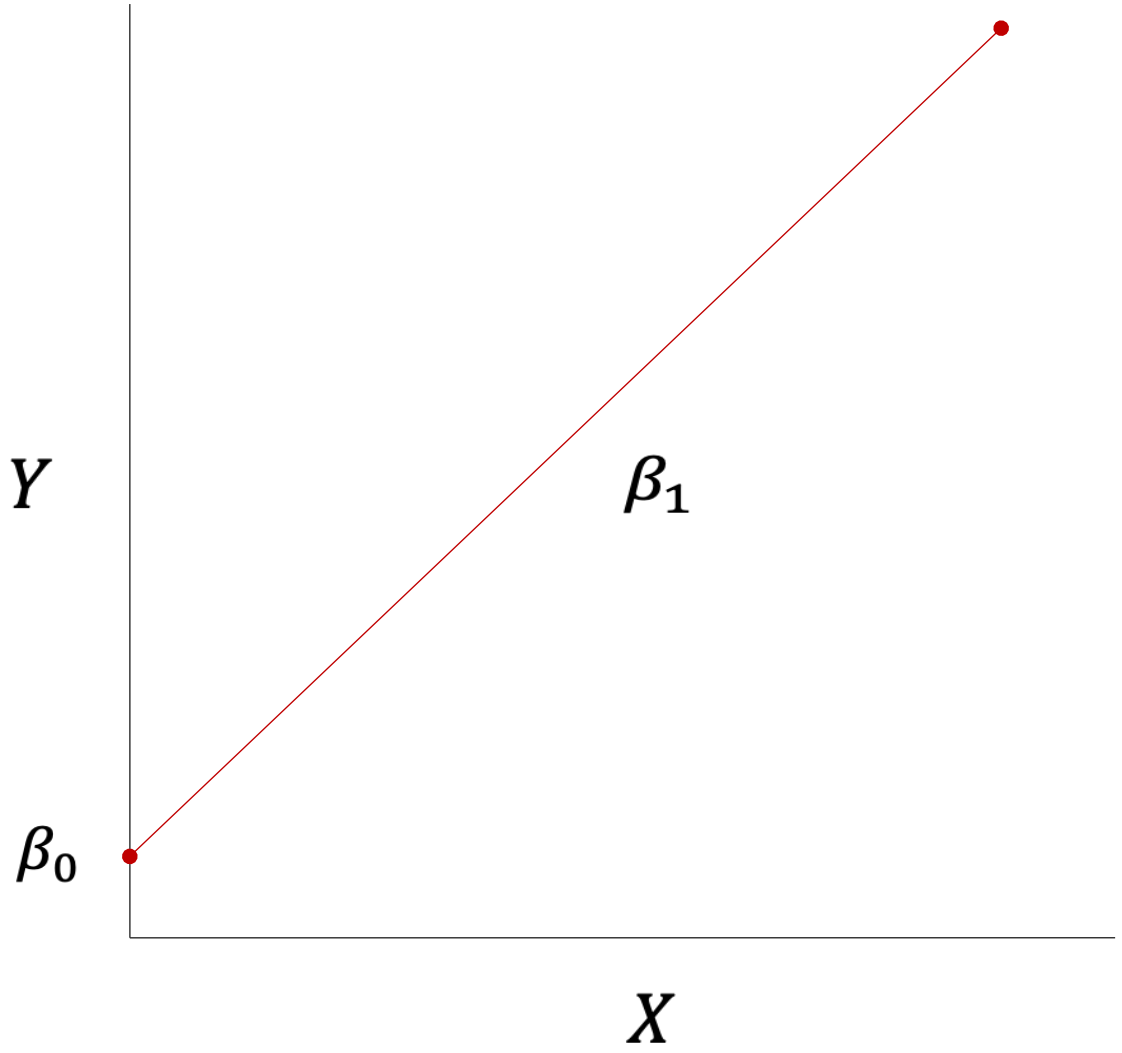


Linear mixed models

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

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.

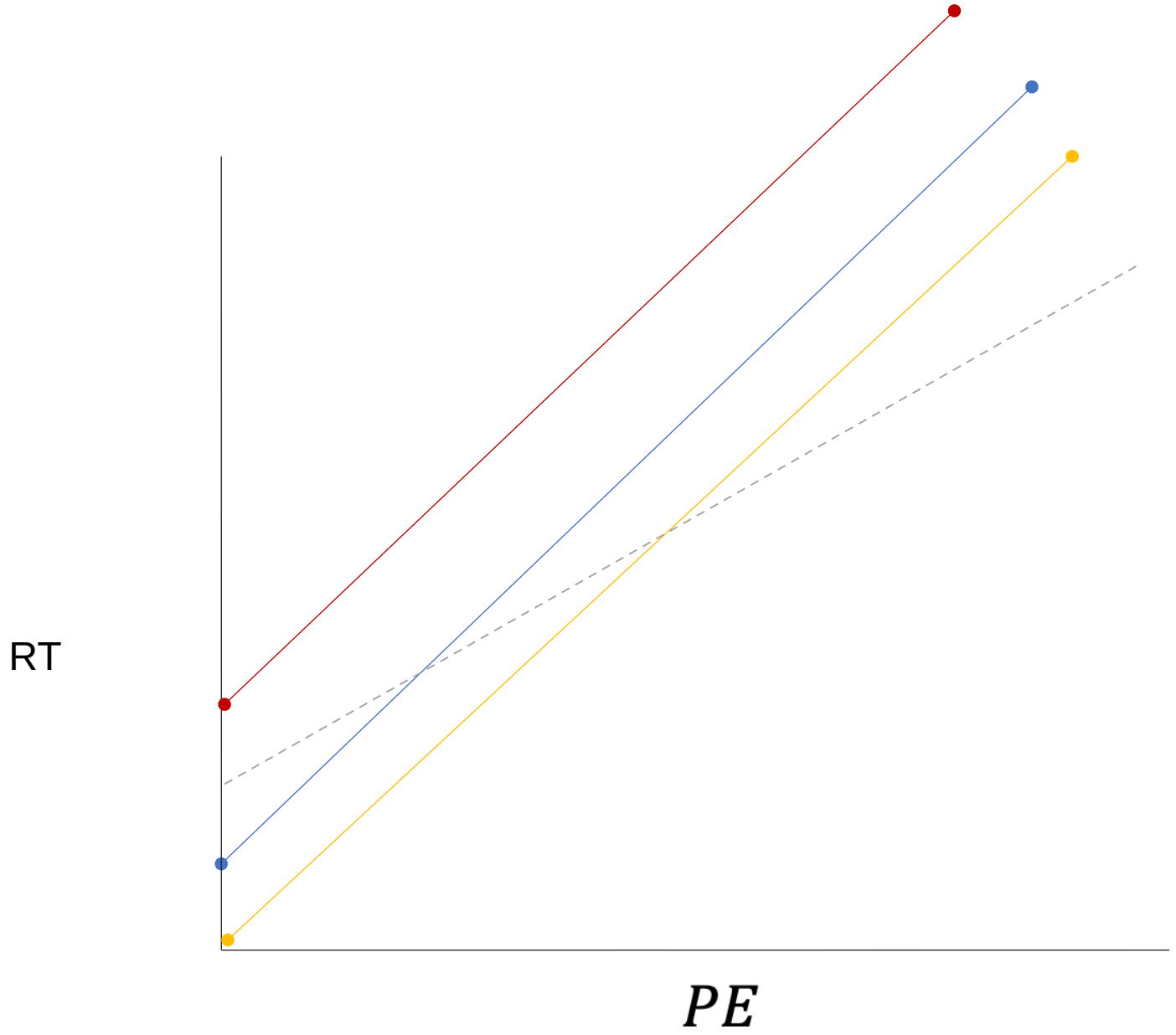
That parameter is also random because we treat it is as being randomly selected from the larger population (Clark, 1973).



Random Intercepts

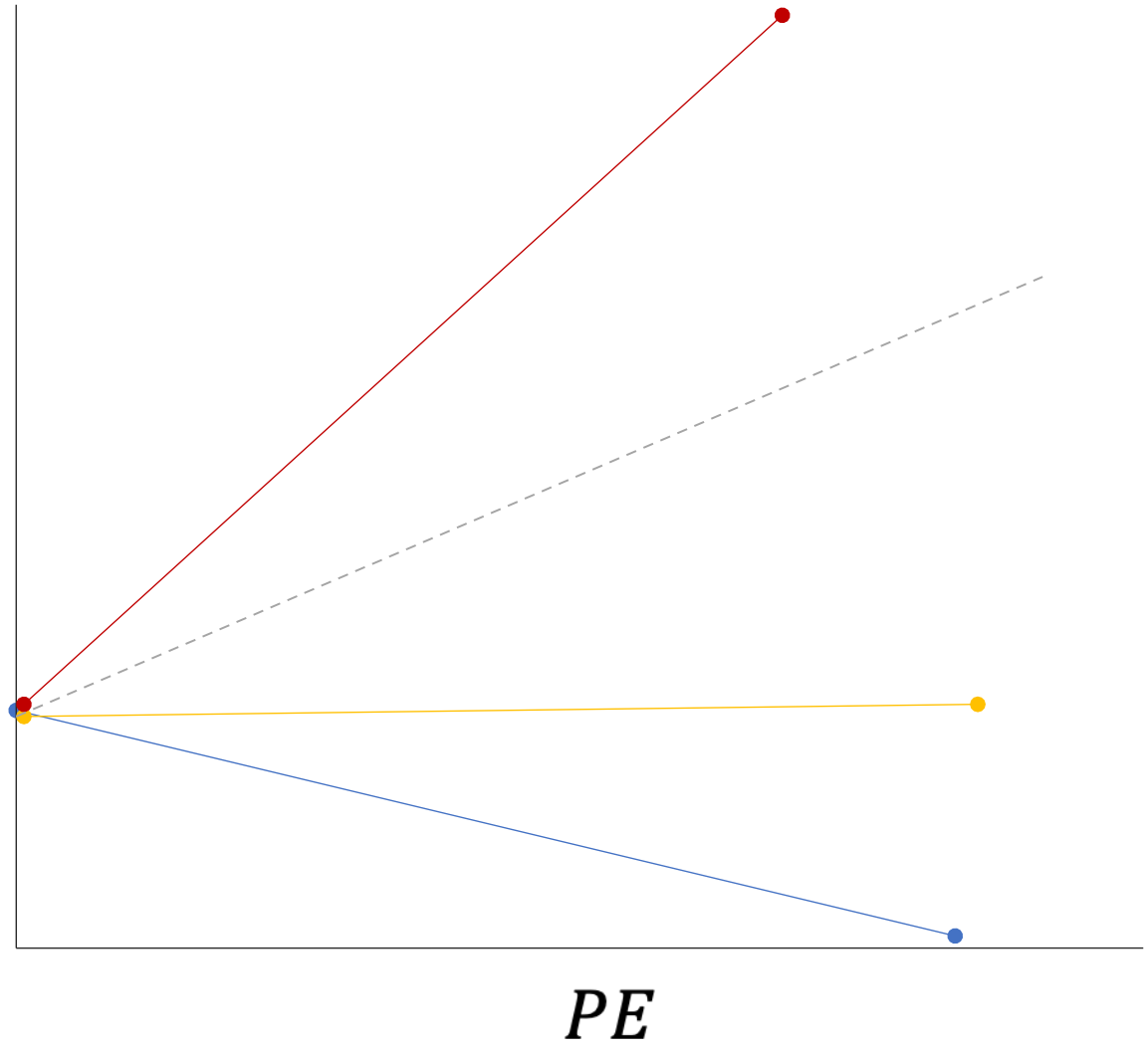
Subject 1
Subject 2
Subject 3

(i.e., differences in “overall” RT)



Random Slopes, fixed intercepts

Subject 1
Subject 2
Subject 3



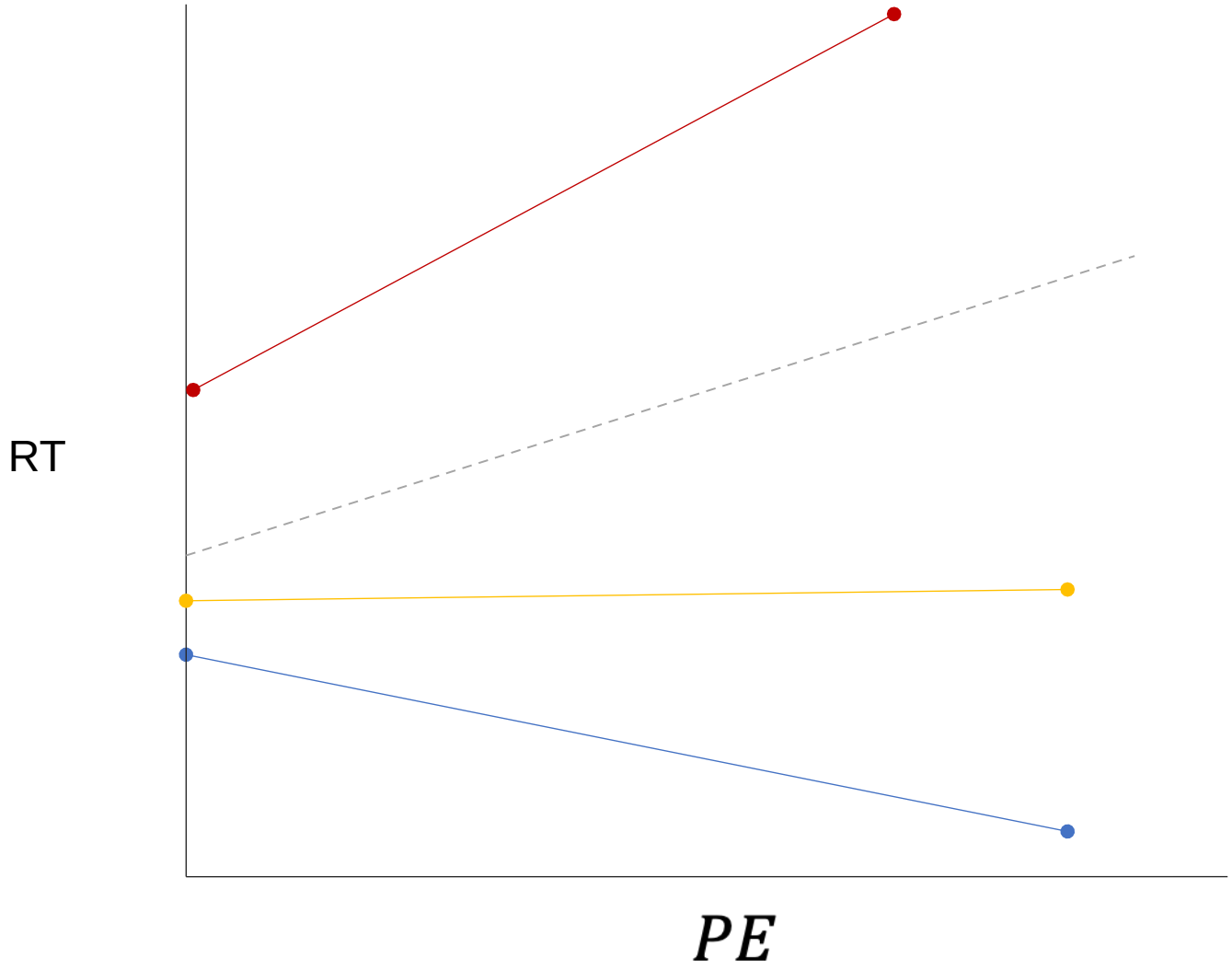
(i.e., differences in the effect across PE levels)

Random Intercepts and Slopes

Subject 1
Subject 2
Subject 3

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

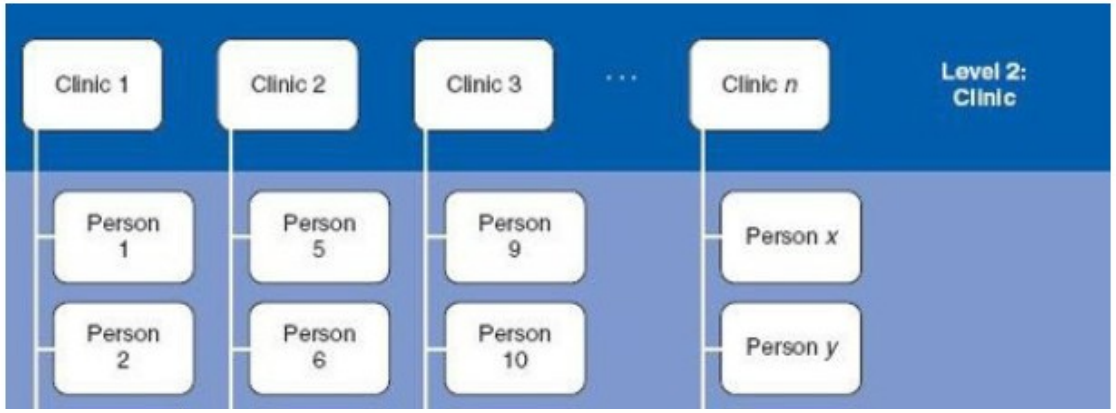
Including random intercepts and slopes allows to control for random variations among items (sampling units)



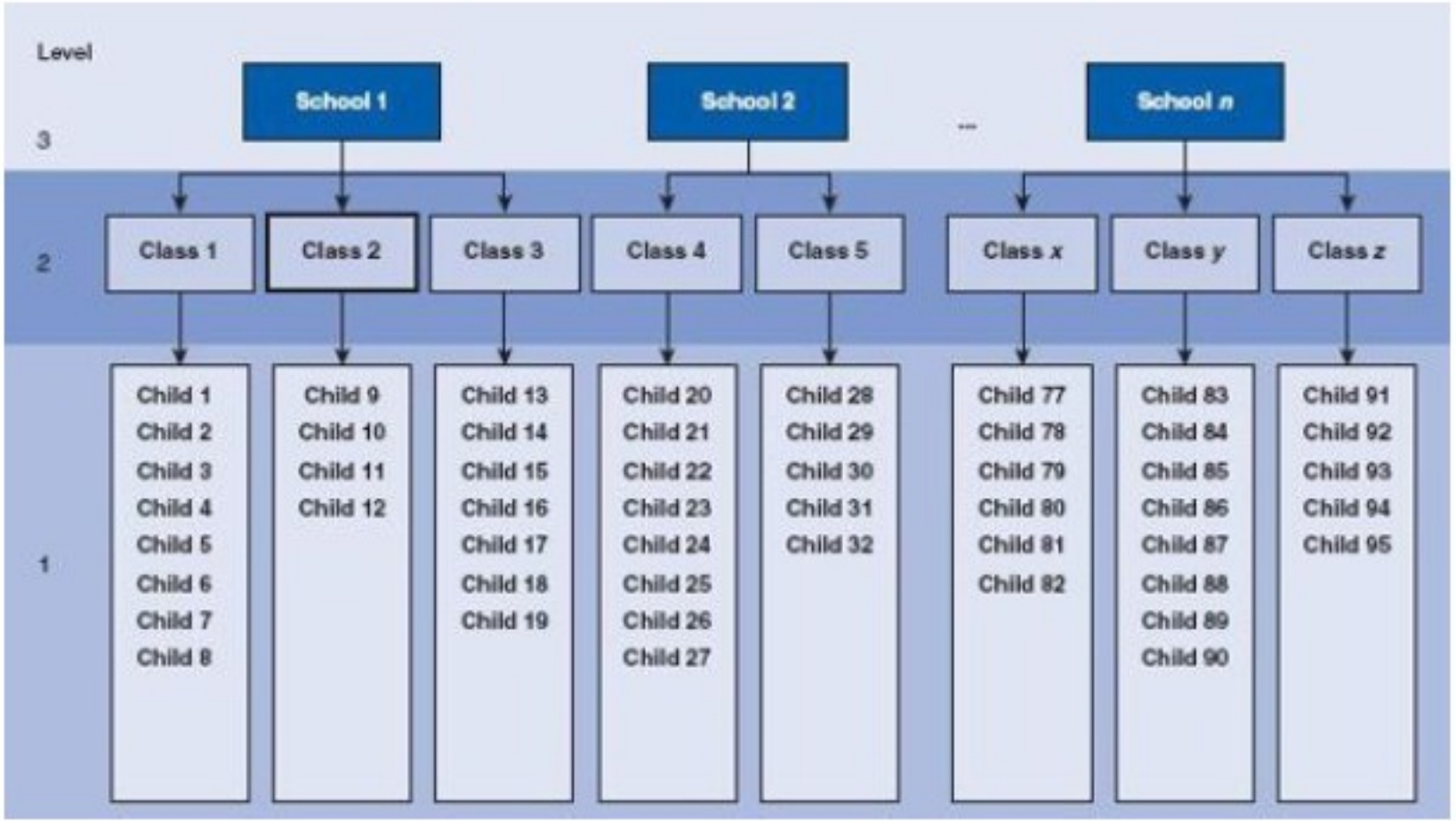
Linear mixed models

Multilevel models

Hierarchical linear mod



Nested data



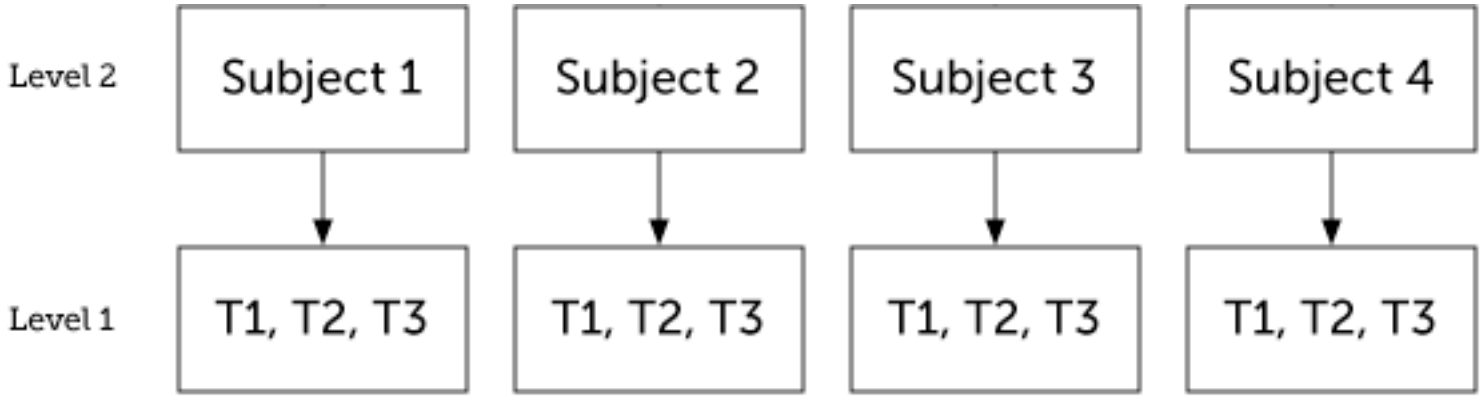
Linear mixed models

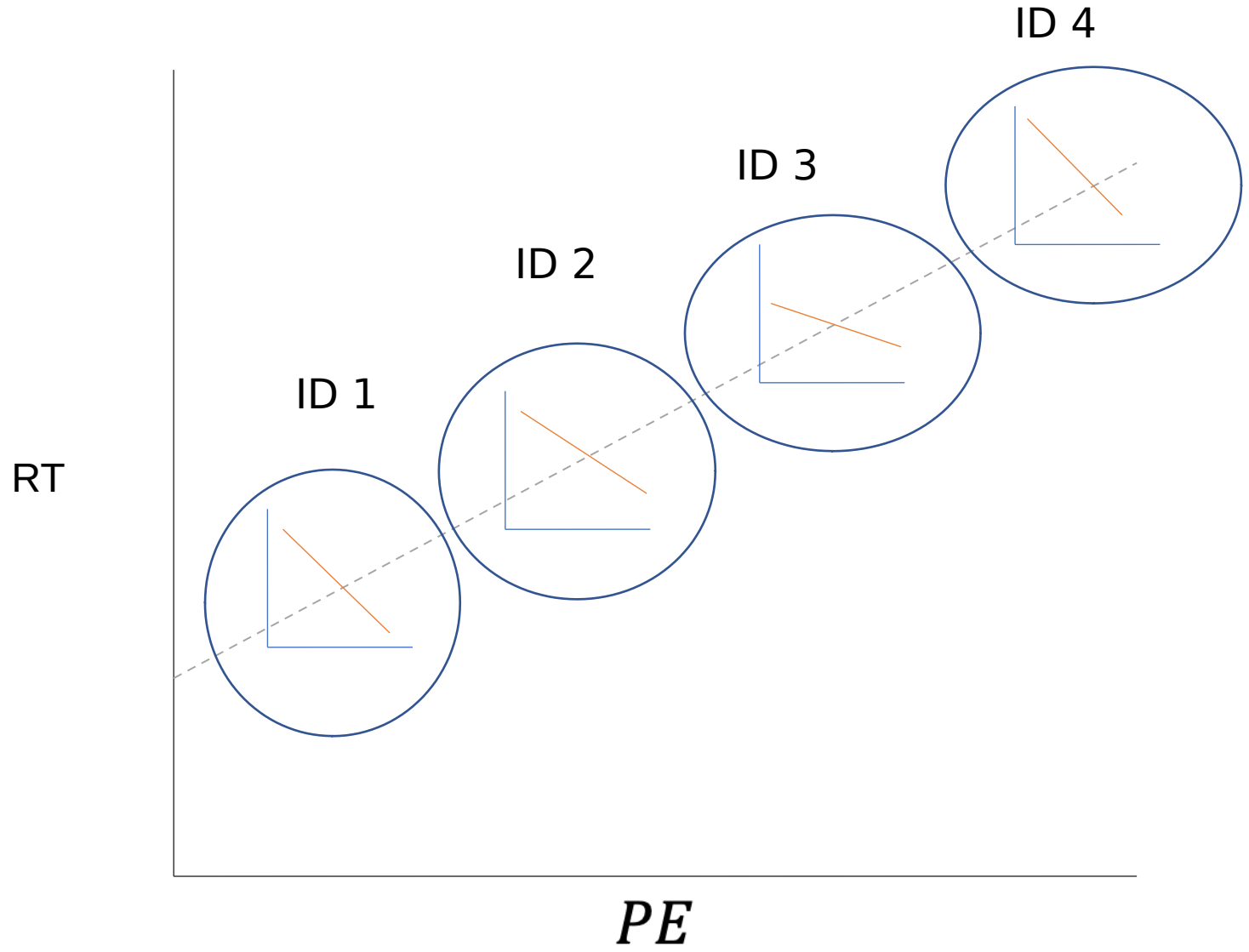
Multilevel models

Hierarchical linear models

Nested data

Taking into account within- and between-person variability

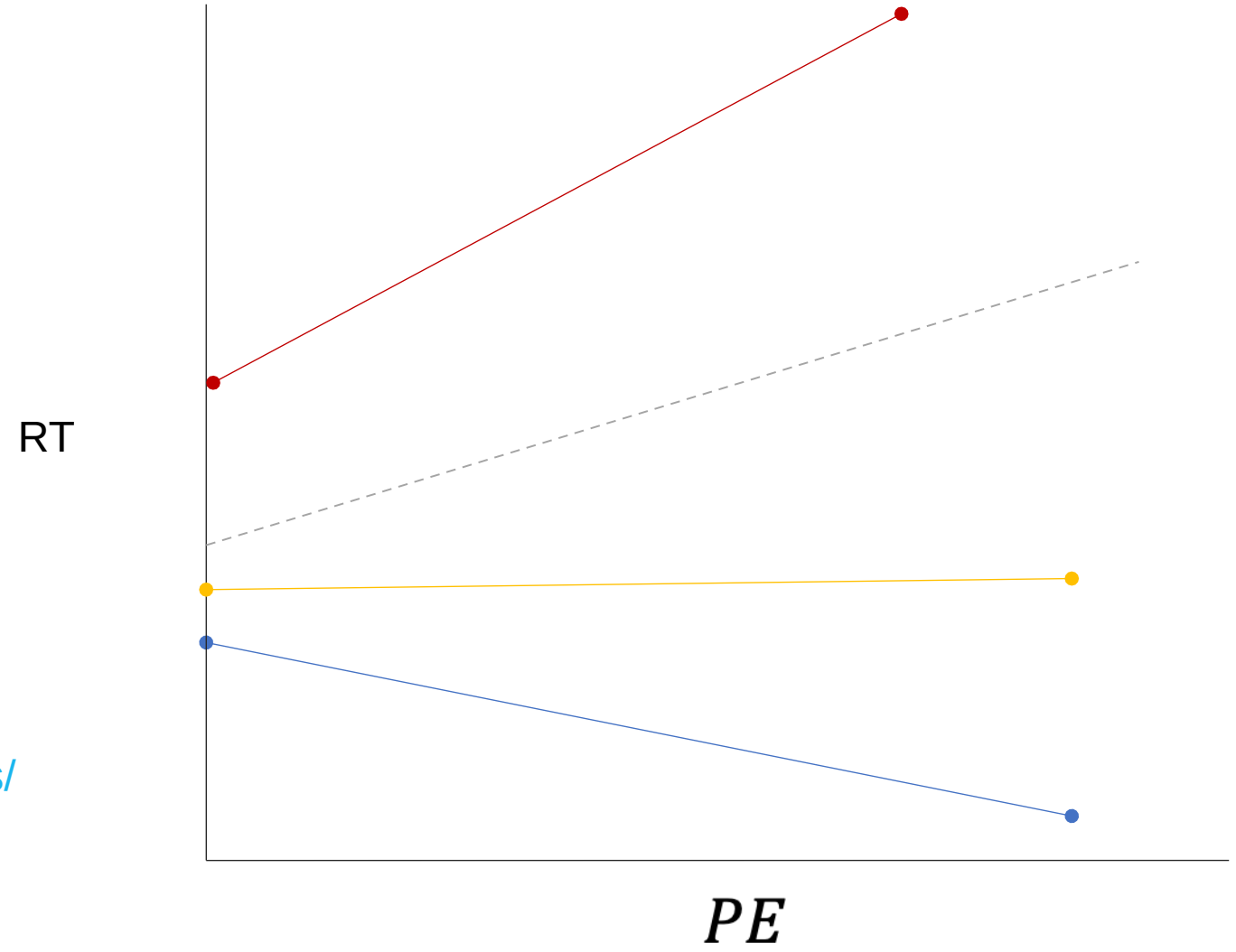




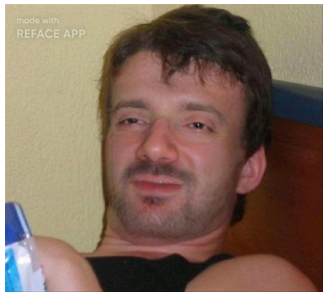
"The reason for favoring the within-subject level is that omitted and confounding variables are less likely to be a problem when analyses focus on how and why people change over time than on how people differ from one another". (Bolger & Laurenceau, *Intensive longitudinal methods: An introduction to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

Random Intercepts and Slopes

Participant 1
Participant 2
Participant 3
Fixed effect (Intercept and slope for the average person)



Check also: http://mfviz.com/hierarchical-models/



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

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 error term $\varepsilon_{ij.}$

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

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

Level 2
(population level)

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

Subject 1

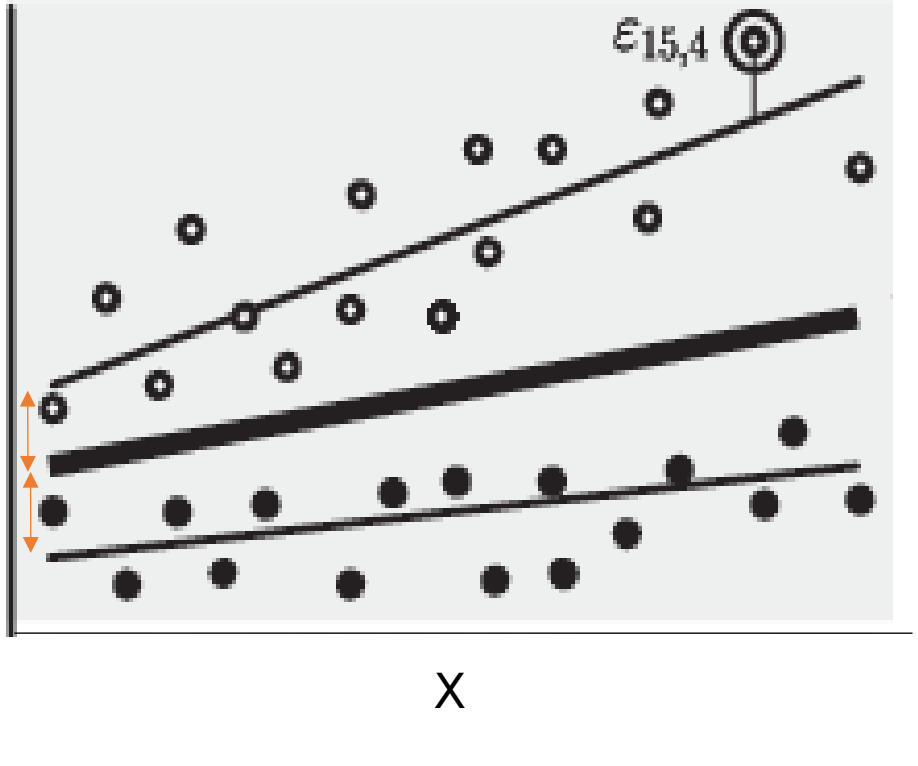
$Y_{i1} = (\gamma_{00} + u_{01})$

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
(Population level)

Subject 1

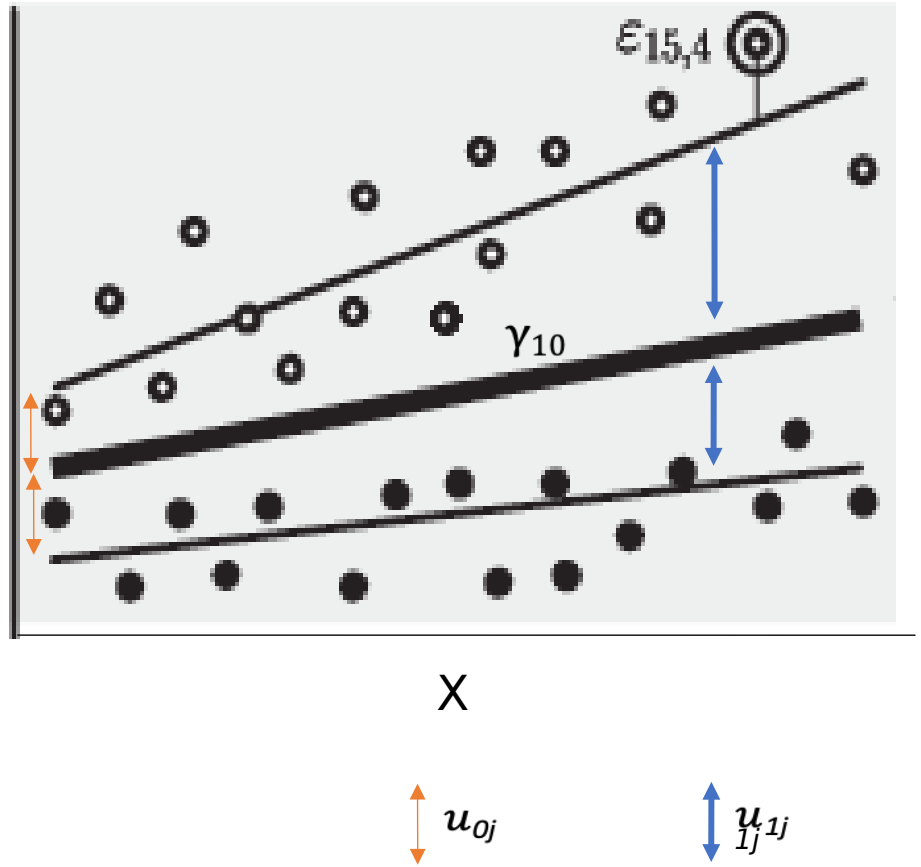
$$Y_{i1} = (\gamma_{00} + u_{01}) + (\gamma_{10} + u_{11}) X_{i1} + \epsilon_{i1}$$

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

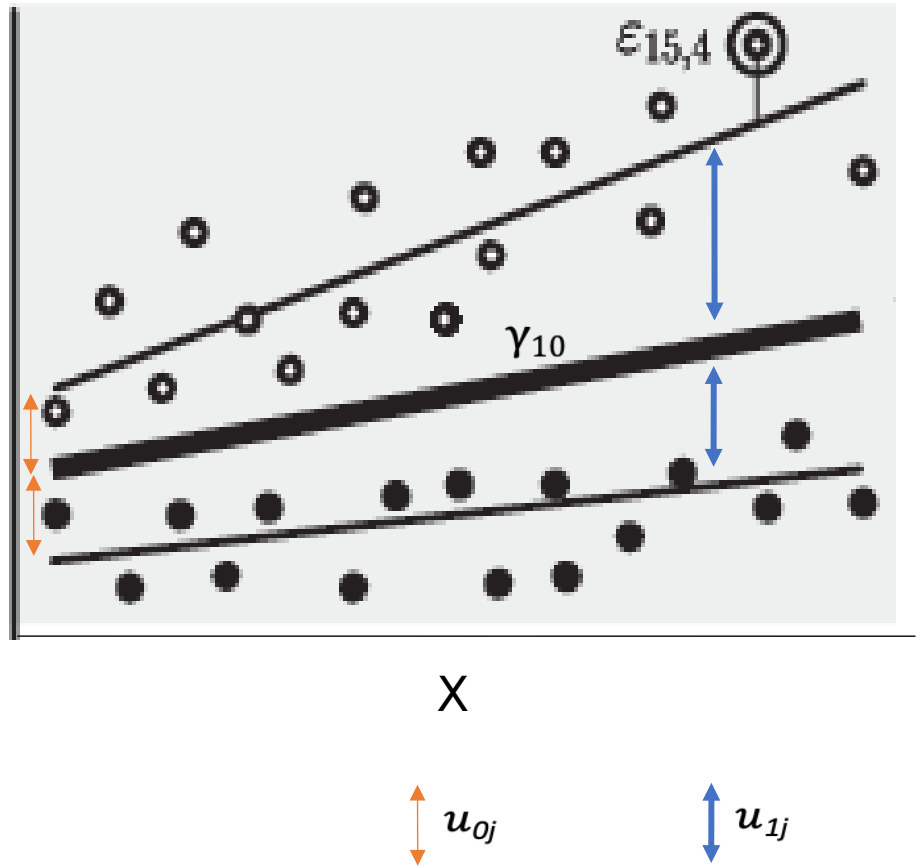
$$Y_{i1} = (\gamma_{00} + u_{01}) + (\gamma_{10} + u_{11}) X_{i1} + \epsilon_{i1}$$

Typical person

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

Subject 2

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



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

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.

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

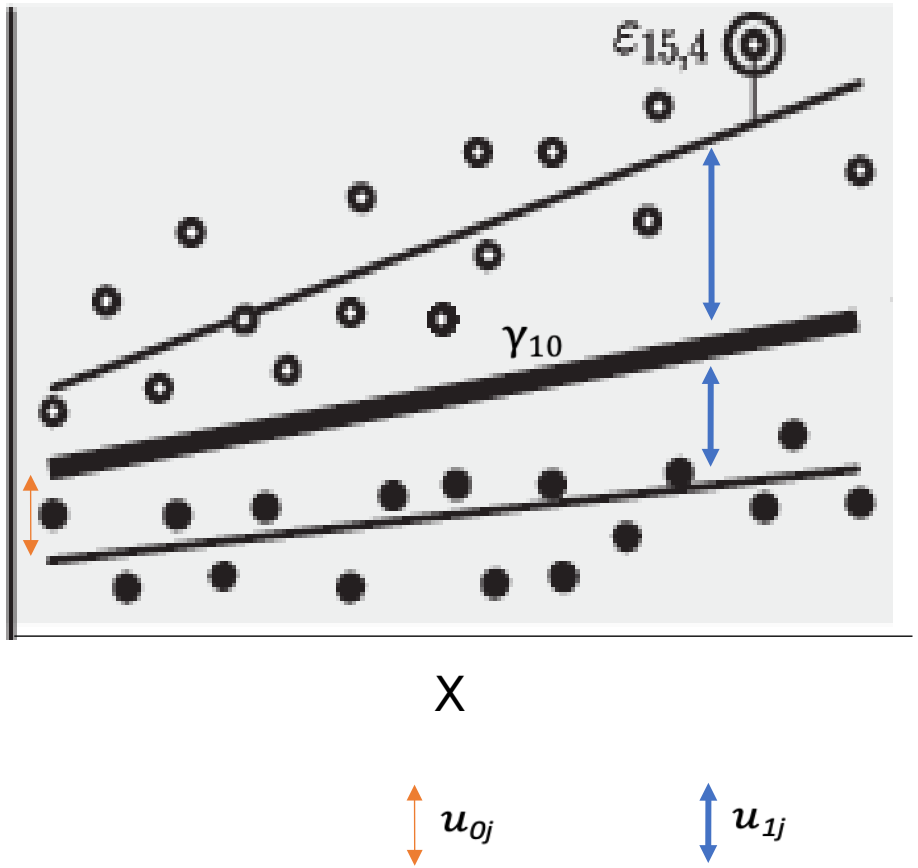
$$\beta_{1j} = \gamma_{10} + u_{1j}$$

Level 2 (between-subjects)

$$u_{0j} \sim N(0, \tau_0^2)$$

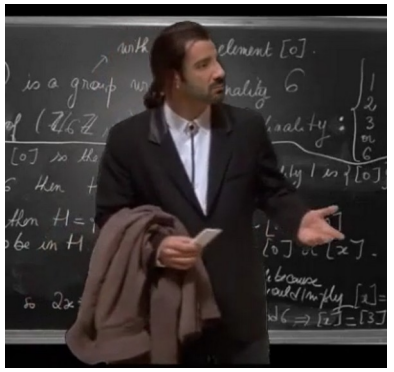
$$u_{1j} \sim N(0, \tau_1^2)$$

$$\epsilon_{1j} \sim N(0, \sigma^2)$$



Model formulation

$$Y_{ij} = \beta_{0j} + \beta_{1j} PE_{ij} + \epsilon_{ij}$$ **Level 1**



$$\beta_{0j} = \gamma_{00} + u_{0j}$$ **Level 2**

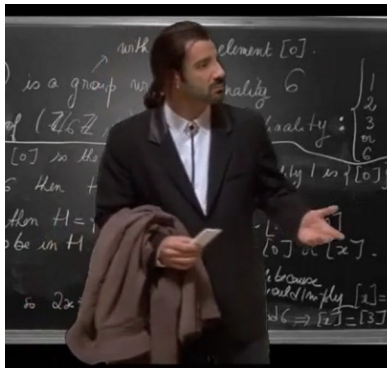
$$\beta_{1j} = \gamma_{10} + u_{1j}$$

$$Y_{ij} = (\gamma_{00} + u_{0j}) + (\gamma_{01} + u_{1j}) PE_{ij} + \epsilon_{ij}$$

If we set the *u*s to 0, we have a fixed effects model

Model formulation

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



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

$$\beta_{1j} = \gamma_{10} + u_{1j}$$ **Level 2**

$$\begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{00}^2 & \tau_{01}^2 \\ \tau_{01}^2 & \tau_{10}^2 \end{pmatrix} \right)$$ **Variance-covariance matrix**

$$\epsilon_{1j} \sim N(0, \sigma^2)$$ **Residuals**

Variance components

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 dataset.
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.

What exactly is Maximum Lixelihood? Watch this:
https://towardsdatascience.com/probability-concepts-explained-maximum-likelihood-estimation-c7b4342fdbb1

Library = lme4

lmm<-lmer(DV~IV+(1+randomeffect|randomintercept),data=df)

Fixed part of the model

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.

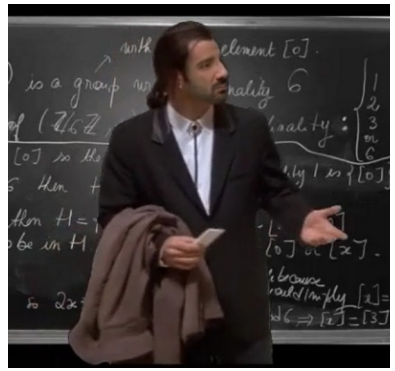
First, let's fit an unconditional model, a model with only the intercept, to examine the variability in our dependent variable that is between- and within- participant

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

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)

τ^2_{0j}

σ^2_{1j}

Random effects:

Groups	Name	Variance	Std.Dev.
subj_id	(Intercept)	3.493	1.869
Residual		4.030	2.008

Number of obs: 9000, groups: subj_id, 30

Lmmworkshop.Rmd

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

$$\rho_I = \frac{\tau^2}{\tau^2 + \sigma^2} \qquad \hat{\rho} = \frac{3.49}{3.49 + 4.03} = 0.46$$

46% of the variance in RT is explained by the grouping variable, “Subjects”.

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

There are several model selection criteria: the most popular ones are:

- Akaine information criterion (AIC)
- Bayes or Schwarz information criterion (BIC)
- Likelihood ratio test

The BIC is known to put a strong penalty on the model complexity for small sample sizes (e.g.,

The anova() function in R allows to carry out a Likelihood ratio test.
In this test, the fit of the full and reduced models are compared.

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

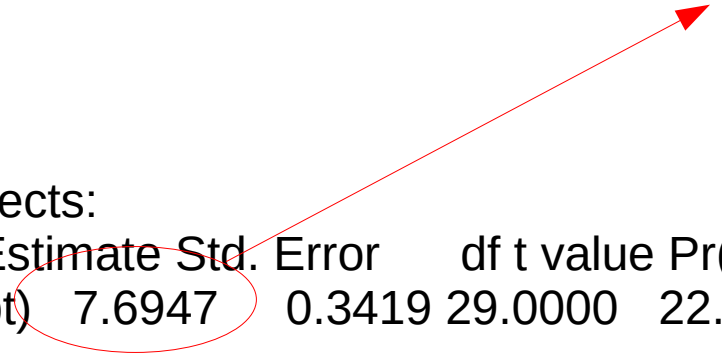
Lmmworkshop.Rmd
```{r lmm rand test sig}
```

```
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 43566 43580 -21781    43562
mixmod_unc    3 38258 38279 -19126    38252 5310.3  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.

Average reaction time



Fixed effects:

	Estimate	Std. Error	df	t value	Pr(>|t|)	
(Intercept)	7.6947	0.3419	29.0000	22.51	<2e-16	***

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}
```

Model selection
At this point in time, there is no perfect solution.

1. Keep it maximal: The model should include the maximal random effects structure *justified by the design* (Barr et al., 2013)

Pro: Avoid the risk of removing variance components (random effects), reducing the possibility of Type I error

Cons: - The most complex model will always provide the best fit for a given data set; However, there is a risk of ovefitting when model are too complex.
- Reduction in power.

2. Balance: The random effect structure should thus be supported by the data (Matuschek et al., 2017)

Barr et al. (2013): “[T]he maximal random effect structure should be fitted to the data. ...The random effect structure should be reduced if and only if the maximal model does not converge.”

Let’s run the maximal model!

```
Library(”lme4”)
maxMod<-lmer(RT~PE+(PE|subj_id), data=data)
```

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

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 relation to the average RT. 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 PE.

Lmmworkshop.Rmd

```{r maximal}
```

Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']
Formula: RT ~ PE + (1 + PE | subj_id)
   Data: df
Control: lmerControl(optimizer = "bobyqa")

REML criterion at convergence: 4179.7

Scaled residuals:
    Min   1Q Median   3Q   Max
-3.3829 -0.6809 0.0004 0.6715 3.6142

Random effects:
 Groups   Name   Variance Std.Dev. Corr
 subj_id (Intercept) 2.50371 1.5823
          PE   12.62806 3.5536   0.44
 Residual   0.08771 0.2962
Number of obs: 9000, groups: subj_id, 30

τ_{00}^2
Variation In RT across subjects

Lmmworkshop.Rmd

```{r maximal}
```

Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']
Formula: RT ~ PE + (1 + PE | subj_id)
   Data: df
Control: lmerControl(optimizer = "bobyqa")

REML criterion at convergence: 4179.7

Scaled residuals:
    Min   1Q Median   3Q   Max
-3.3829 -0.6809 0.0004 0.6715 3.6142

Random effects:
 Groups   Name   Variance Std.Dev. Corr
 subj_id (Intercept) 2.50371 1.5823
          PE   12.62806 3.5536   0.44
 Residual   0.08771 0.2962
Number of obs: 9000, groups: subj_id, 30

τ_{10}^2
Variation in the effect on PE on RT across subjects

Lmmworkshop.Rmd

```{r maximal}
```

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']
Formula: RT ~ PE + (1 + PE | subj_id)
   Data: df
Control: lmerControl(optimizer = "bobyqa")

REML criterion at convergence: 4179.7

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.3829 -0.6809  0.0004  0.6715  3.6142

Random effects:
 Groups   Name        Variance Std.Dev. Corr
 subj_id  (Intercept)  2.50371 1.5823
          PE                   12.62806 3.5536   0.44
 Residual                0.08771 0.2962
Number of obs: 9000, groups:  subj_id, 30
```

τ_{01}^2

Correlation betwee randon intercepts and random slopes

Lmmworkshop.Rmd

```{r maximal}
```

γ_{10}

Fixed effect of PE

Fixed effects:

	Estimate	Std. Error	df	t value	Pr(>|t|)	
(Intercept)	7.4325	0.2889	28.9999	25.726	< 2e-16	***
PE	1.8175	0.6488	29.0001	2.801	0.00897	**

We could also have included between-participant effects in the same model,
By using mean centering

Person-mean centering: subtracing the grand mean from the person mean

Grand-mean centering: subtracting the person mean from the grand mean

Lmmworkshop.Rmd

```{r centering}
```

Pseudo R^2

How much variance does our predictor (predictors) explain on level 1 and level 2?

•Can be calculated for multiple UVs (in this example only PE as UV)

Lmmworkshop.Rmd
```{r Calculating R2pseudo between, within and between-within}
```

R^2pseudo:within

•Quantify change in residual variance (within, level 1)

$$R^2_{\text{pseudo:within}} = \frac{\mathbb{V}ar[\varepsilon] - \mathbb{V}ar[\varepsilon^*]}{\mathbb{V}ar[\varepsilon]} = 1 - \frac{\mathbb{V}ar[\varepsilon^*]}{\mathbb{V}ar[\varepsilon]}$$

Residual variance mixmod_unc

Residual variance ModPE

Models:

•ModPE<-lmer(RT~PE.gmc+(1|subj_id), data = df, control=lmerControl(optimizer="bobyqa"))

•mixmod_unc<-lmer(RT~1+(1|subj_id), data = df)

Pseudo R^2

How much variance does our predictor (predictors) explain on level 1 and level 2?

•Can be calculated for multiple UVs (in this example only PE as UV)

R^2pseudo:between

•Quantify change in intercept variance (between, level 2)

$$R^2_{\text{pseudo:between}} = \frac{\mathbb{V}ar[u_0] - \mathbb{V}ar[u_0^*]}{\mathbb{V}ar[u_0]} = 1 - \frac{\mathbb{V}ar[u_0^*]}{\mathbb{V}ar[u_0]}$$

Intercept variance mixmod_unc

Intercept variance ModPE

Lmmworkshop.Rmd

```{r Calculating R2pseudo between, within and between-within}
```

Models:

•ModPE<-lmer(RT~PE.gmc+(1|subj_id), data = df, control=lmerControl(optimizer="bobyqa"))

•mixmod_unc<-lmer(RT~1+(1|subj_id), data = df)

Pseudo R^2

How much variance does our predictor (predictors) explain on level 1 and level 2?

•Can be calculated for multiple UVs (in this example only PE as UV)

R^2pseudo:bw (between and within)

•How much variance of criterion can be explained by predictor

$$R^2_{\text{pseudo:bw}} = \frac{(\mathbb{V}ar[u_0] + \mathbb{V}ar[\varepsilon]) - (\mathbb{V}ar[u_0^*] + \mathbb{V}ar[\varepsilon^*])}{\mathbb{V}ar[u_0] + \mathbb{V}ar[\varepsilon]}$$
$$= 1 - \frac{\mathbb{V}ar[u_0^*] + \mathbb{V}ar[\varepsilon^*]}{\mathbb{V}ar[u_0] + \mathbb{V}ar[\varepsilon]}.$$

•Explained variances of ModPE
•and mixmod_unc are compared

Lmmworkshop.Rmd
```{r Calculating R2pseudo between, within and between-within}
```

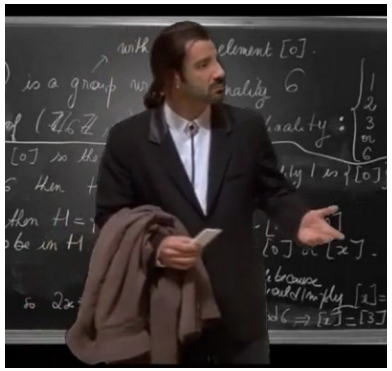
Models:

•ModPE<-lmer(RT~PE.gmc+(1|subj_id), data = df, control=lmerControl(optimizer="bobyqa"))

•mixmod_unc<-lmer(RT~1+(1|subj_id), data = df)

Model formulation

```
maxMod<-lmer(RT~PE+(PE|subj_id), data=data)
```



$$\beta_{ij} = \beta_{0j} + \beta_{1j} PE_{ij} + \epsilon_{ij}$$

Level 1

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

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

Level 2

$$\begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{00}^2 & \tau_{01}^2 \\ \tau_{01}^2 & \tau_{10}^2 \end{pmatrix} \right)$$

Variance-covariance matrix - unstructured

$$\epsilon_{1j} \sim N(0, \sigma^2)$$

Residuals

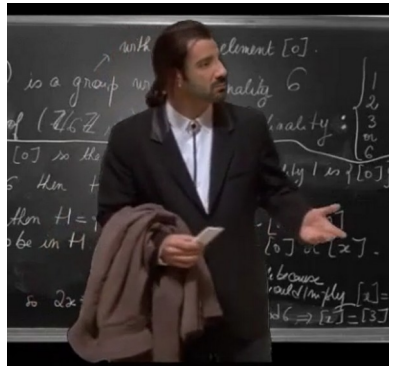
Variance-covariance structures

$$\begin{pmatrix} \tau_{00}^2 & \tau_{01}^2 \\ \tau_{01}^2 & \tau_{10}^2 \end{pmatrix}$$

Unstructured: the covariances can take any value (finite positive)

$$\begin{pmatrix} \tau_{00}^2 & 0 \\ 0 & \tau_{10}^2 \end{pmatrix}$$

Diagonal: the variances are assumed to be indipendent, so their covariance is set at 0



Model Selection

You can start but simple models and add random effects one by one – Forward selection

Or start from the maximal model and take out random effects – Backward selection

Both approaches are okay, as look as you add a clear statement of the critera used when selecting both fixed and random effects (Meteyard & Davies, 2020).

Model Comparison

Matuschek et al. (2017) : Backward selection.
Start from the most complex model (maximal model) until a further reduction would imply a significant loss in the goodness of fit.

Let’s try it!

Let’s fit a model where the covariance between random effect is set to zero.

Lmmworkshop.Rmd
```{r model selection backward}
```

Model Comparison

Matuschek et al. (2017) : we should not set α = 0.05 as cut-off for choosing models, because it may imply a strong penalty on the model complexity, in favour of the goodness-of-fit.

They suggested a less conservative alpha, α = 0.2

Example of backward model selection

Mod0 = lmer(RT~PE+(PE|subj_id), data=data) **maximal**

Mod1 = lmer(RT~PE+(PE||subj_id), data=data) **Drop RE covariance**

Mod2 = lmer(RT~PE+(1|subj_id), data=data) **Drop random slopes**

Model Comparison

Excercise: Forward model selection, testing the random effects of items (crossed effects).

Lmmworkshop.Rmd

```{r  exercise1 model selection forward}
```

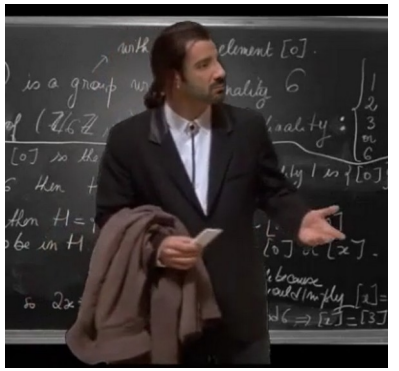
Mod0 = ?

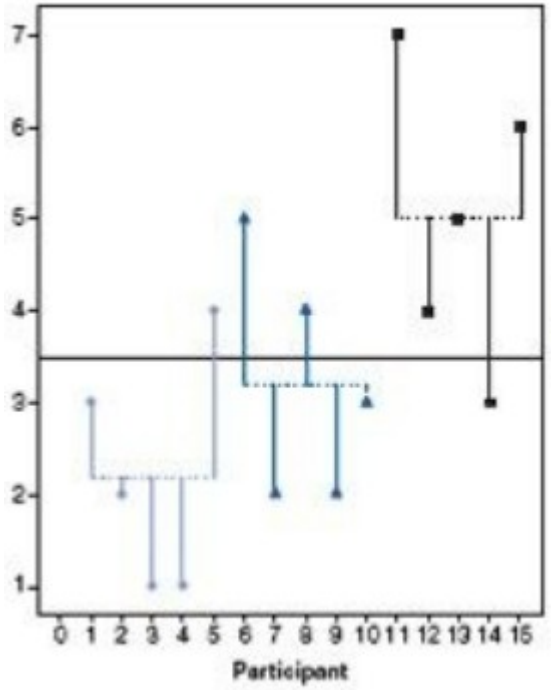
Mod1 = ?

Mod2 = ?

categorical predictor – ANOVA

$$Y_i = \beta_0 + \beta_1 PElow + \beta_2 PEhigh + \epsilon_i$$



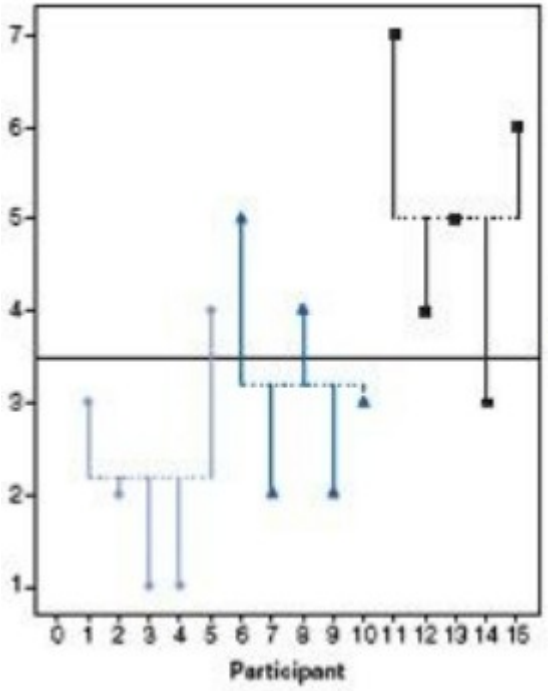


categorical predictor – ANOVA

$$Y_i = \beta_0 + \beta_1 PElow + \beta_2 PEhigh + \epsilon_i$$

Lmmworkshop.Rmd

```{r categorical predictor between}
```



categorical predictor – between participants

$$Y_i = \beta_0 + \beta_1 PElow + \beta_2 PEhigh + \epsilon_i$$

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	7.380176	0.595864	12.386	1.2e-12 ***
PEbwHighPE	0.937268	0.842680	1.112	0.276
PEbwLowPE	0.006359	0.842680	0.008	0.994

$\beta_0 = 7.38$

$\beta_1 = 7.38 + 0.94 = 8.32$

$\beta_2 = 7.38 + 0.01 = 7.39$

PebwHighPE = PEHigh-PEMedium
PebwLowPE = PELow - PEMedium

categorical predictor – between participants

$$Y_i = \beta_0 + \beta_1 PElow + \beta_2 PEhigh + \epsilon_i$$

We can also set meaningful contrasts

Lmmworkshop.Rmd
```{r setting contrasts}
```

Is it the same if we transform categorical into continuous?

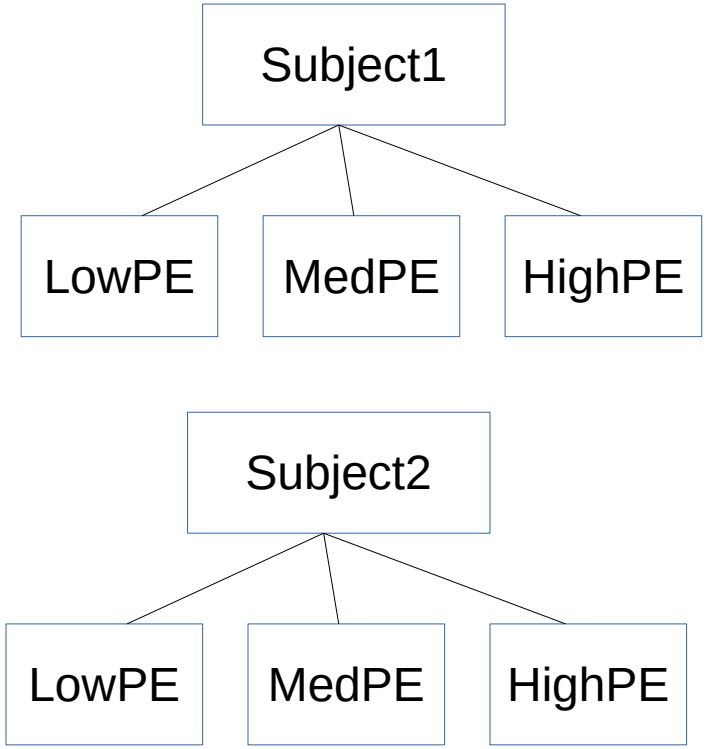
Continuous variable: if we treat a variable as continuous we are willing to treat the differences between values as quantitative. Thus, it is meaningful to talk about the effct of “ a one-point increase” in the values of X.

Categorical (discrete) variable: We treat the values of the variable as discrete, and estimate their effects separately. Thus, the effect of moving from one category to another may differ depending on the categories.

Lmmworkshop.Rmd

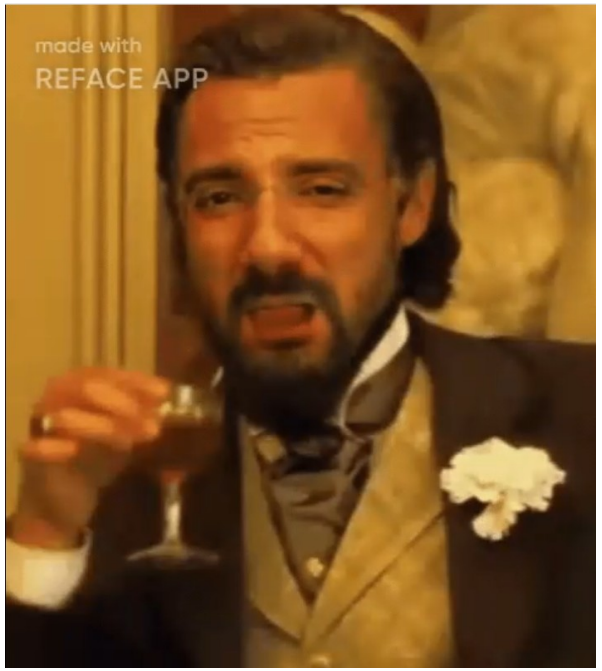
```{r continuous predictor between}
```

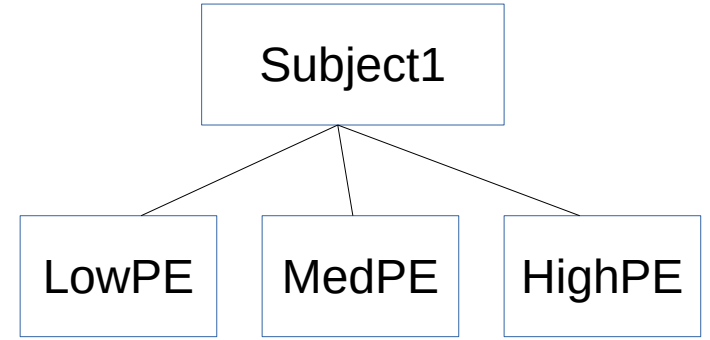
categorical predictor – within participants

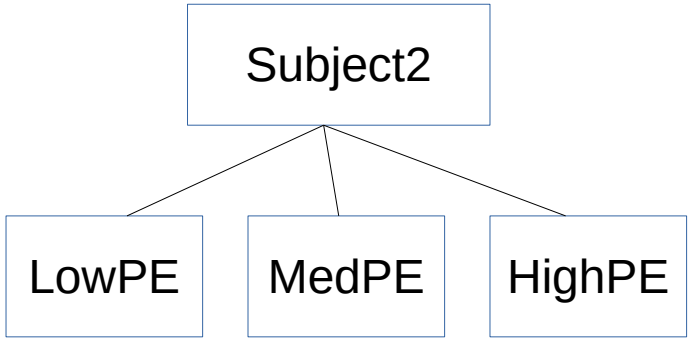


categorical predictor – within participants

Let’s take it EZ!







Lmmworkshop.Rmd

```{r continuous predictor within ez}
```

categorical predictor – within participants, mixed effects

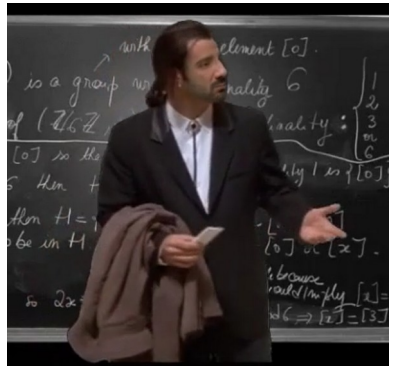
$$Y_{ij} = \beta_{0j} + \beta_{1j} PElow_i + \beta_{2j} PEhigh_i + \epsilon_{ij}$$

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

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

$$\beta_{2j} = \gamma_{20} + u_{2j}$$

$$\begin{pmatrix} u_{0j} \\ u_{1j} \\ u_{2j} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{00}^2 & \tau_{01}^2 & \tau_{02}^2 \\ \tau_{01}^2 & \tau_{10}^2 & \tau_{12}^2 \\ \tau_{02}^2 & \tau_{12}^2 & \tau_{20}^2 \end{pmatrix} \right)$$



categorical predictor – within participants, mixed effects

Lmmworkshop.Rmd

```{r categorical predictor within mixed-effects}
```

lme4 or nlme?

Lme4 tend to be preferred because of some advantages over nlme:

- It is faster
- It implements crossed effects easily
- Allows the creation of GLMMMs
- Cutting-edge

Lmmworkshop.Rmd

```{r  exercise2 contrasts for within lmm}
```

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
- More reliable estimates, even in the presence of outliers

Recommended readings:

- Finch, W. H., Bolin, J. E., & Kelley, K. (2019). Multilevel modeling using R. Crc Press.
- Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. Journal of memory and language, 94, 305-315.
- Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. Behavior research methods, 49(4), 1494-1502.
- Meteyard, L., & Davies, R. A. (2020). Best practice guidance for linear mixed-effects models in psychological science. Journal of Memory and Language, 112, 104092.

Example of reporting:

”Models were fitted (using the lme4 package version 1.1-13, Bates, Mächler, Bolker, & Walker, 2015; R Core Team, 2014)“

”Analyses were run using the lme4 package in R (Bates et al., 2015)”.

”In order to select random effect, a forward approach was used. The selection started with a model without random effects. Then this model was compared with a model were random intercepts were included. Finally, a model with random intercept and slopes was tested. To test the significant of each random effect, we used likelihood ratio test (LRT) for each pairwise comparison, with $\alpha_{LRT} = 0.2$ (Matshucheck et al., 2017)”

”To test fixed effect, models were estimated with REML and Satterthwaite and Kenward-Rogers approximations were applied, as this method has been shown to provide robust estimations (Luke, 2016).