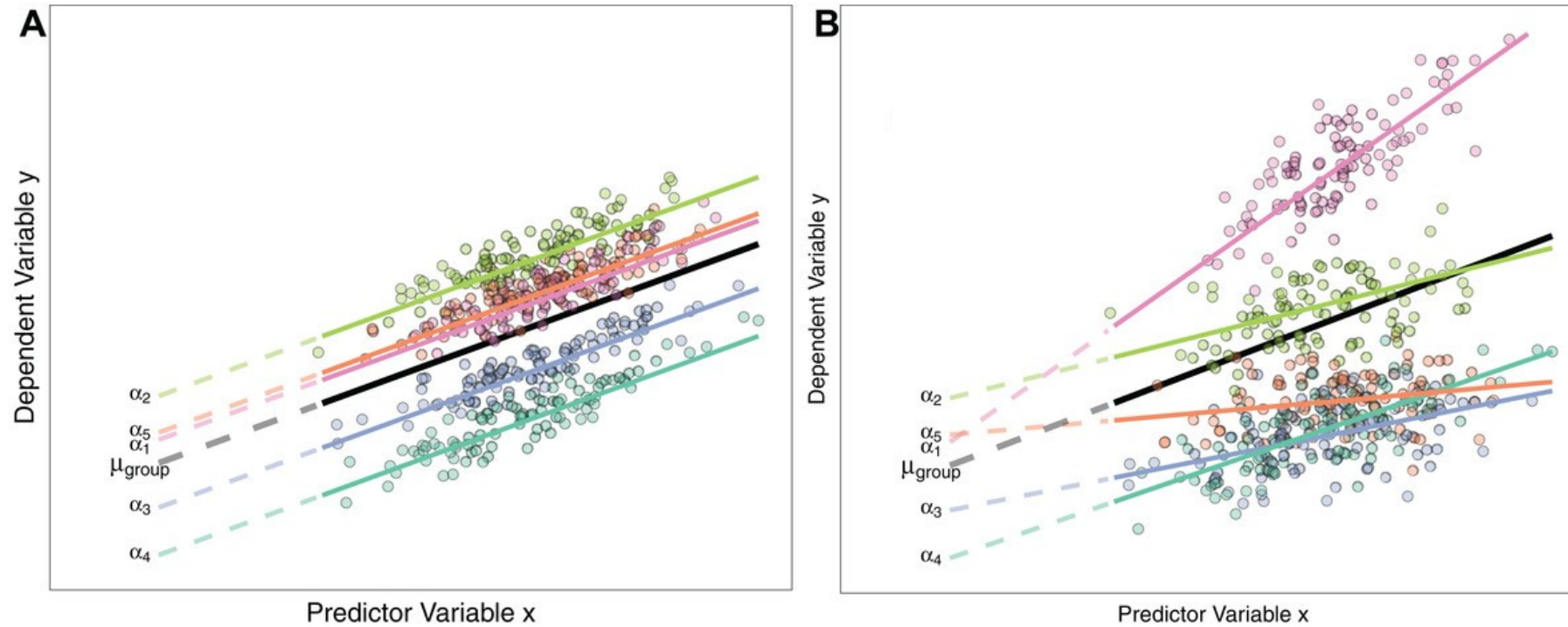


Introduction to:

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 going to cover?

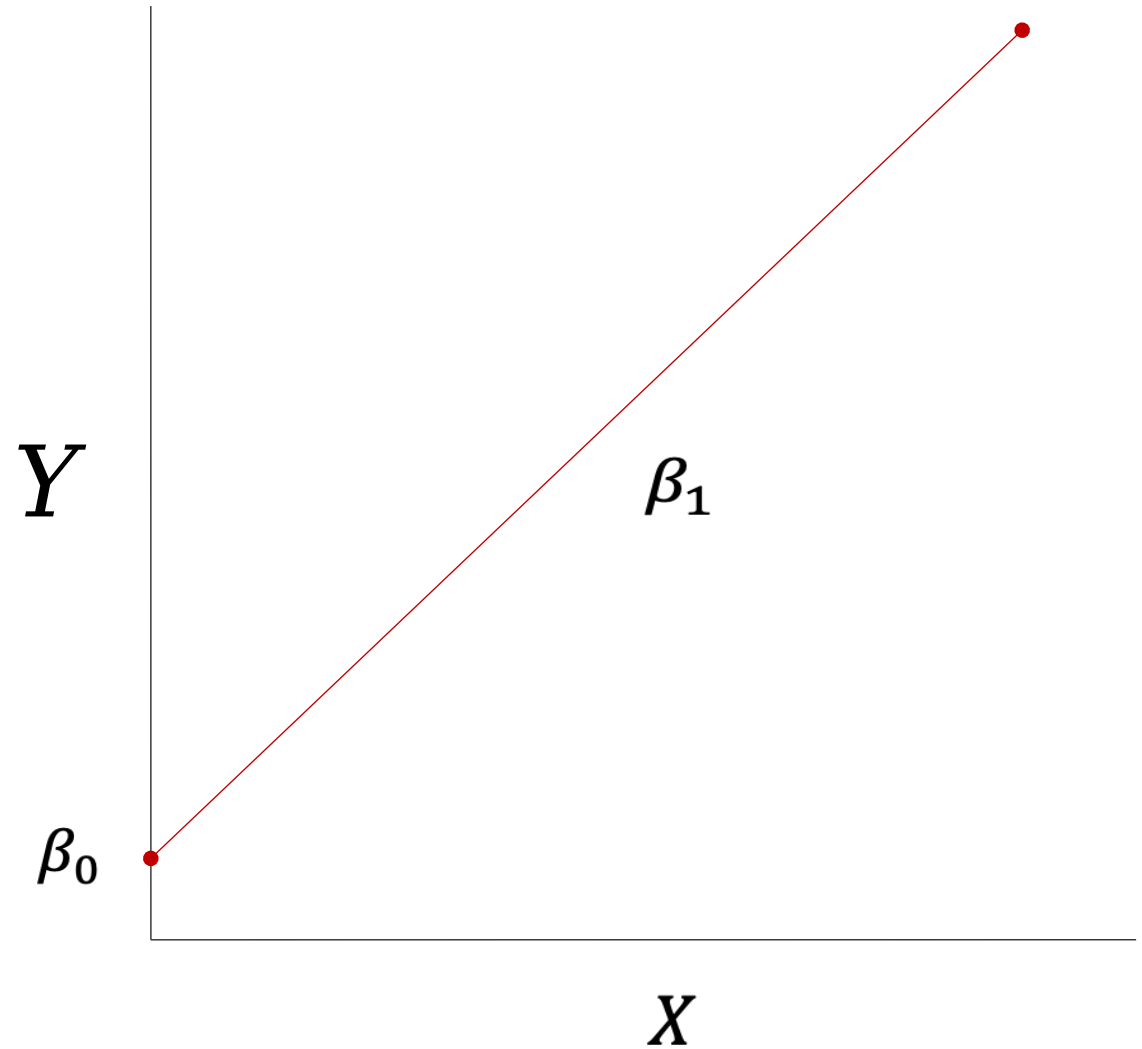
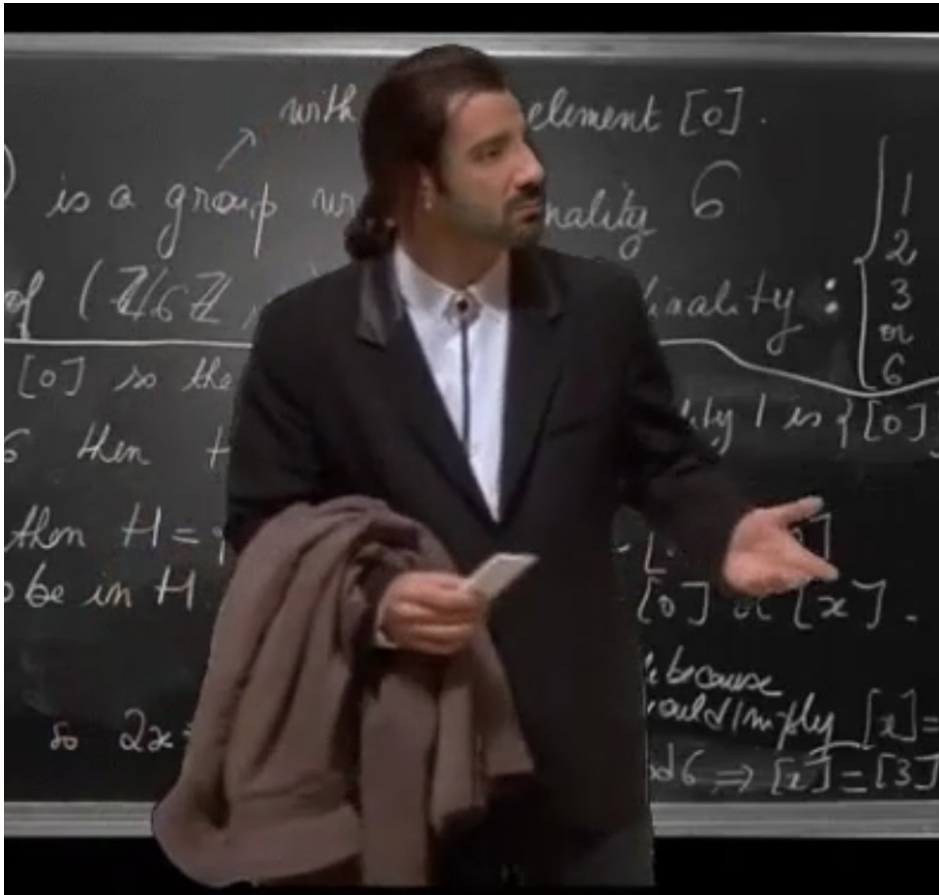
- Linear models: the basics
- Linear mixed - effects models: random effects
- Parameter estimation
- Intraclass correlation
- Calculating R^2 within- between
- Variance - covariance matrices
- Model selection
- Linear models with categorical predictors
- Setting contrasts
- Linear models with binary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

Theory

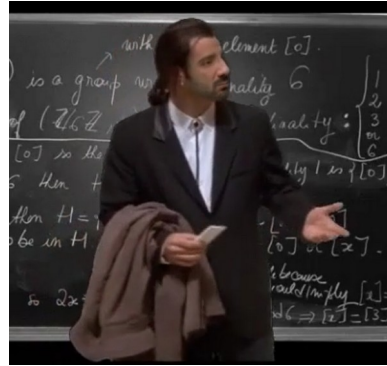
Practice

Exercises

Linear models



Linear models



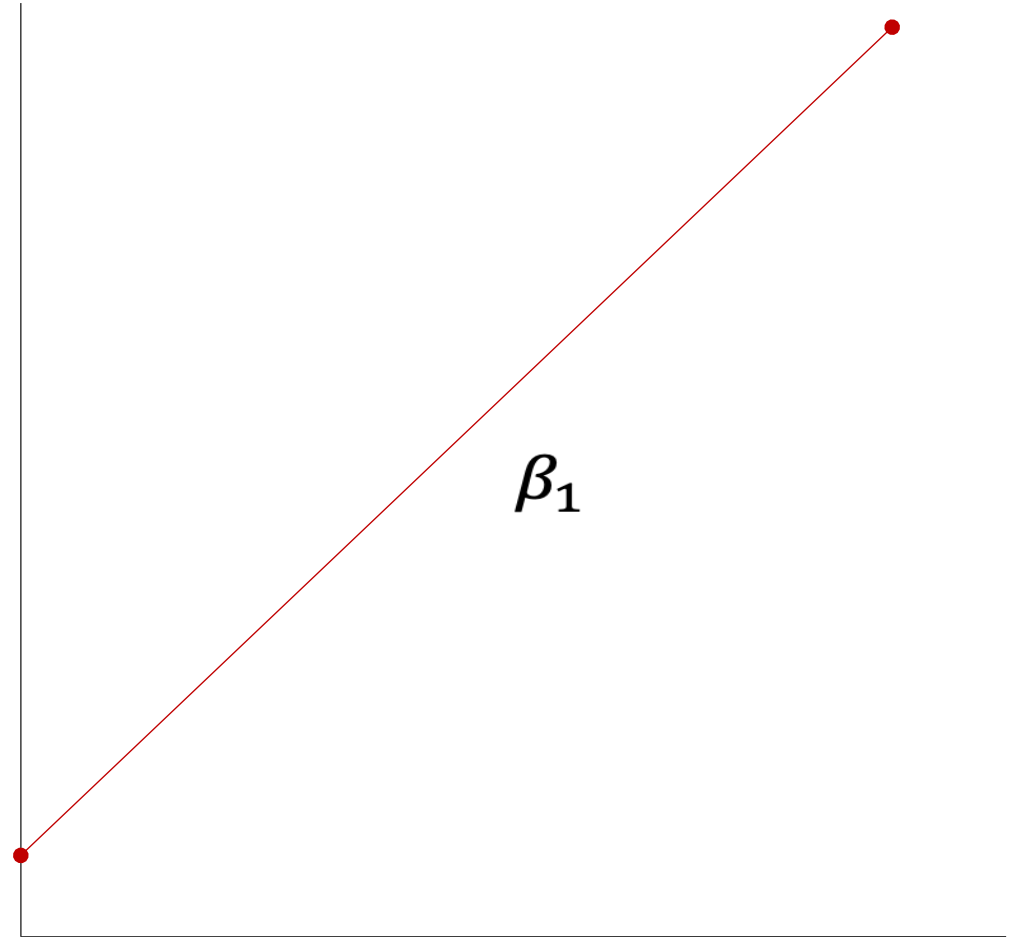
$$Y = \beta_0 + \beta_1 X$$

Y

β_1

β_0

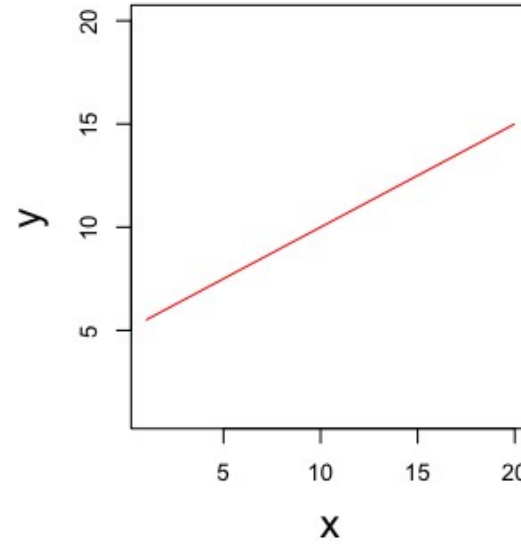
X



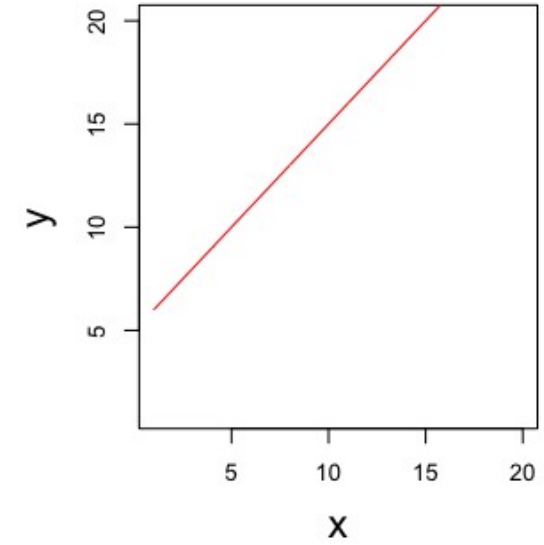
Linear models

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

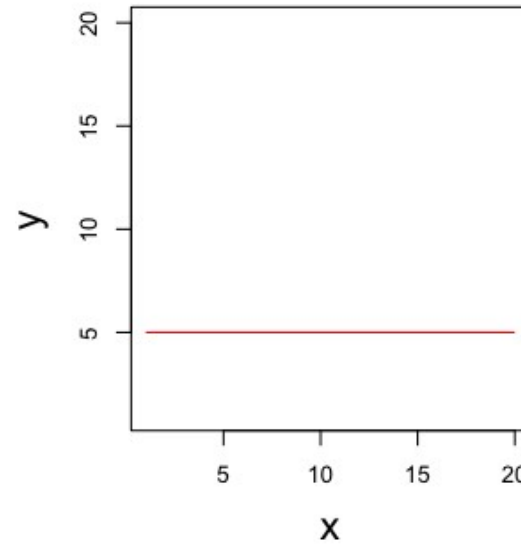
$b_0 = 5, b_1 = 0.5$



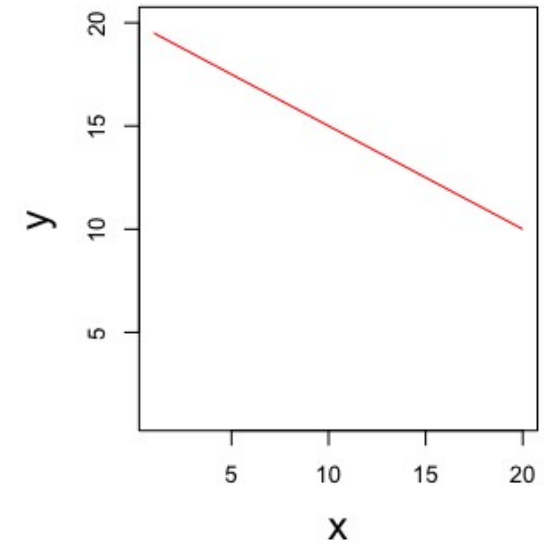
$b_0 = 5, b_1 = 1$



$b_0 = 5, b_1 = 0$



$b_0 = 20, b_1 = -0.5$



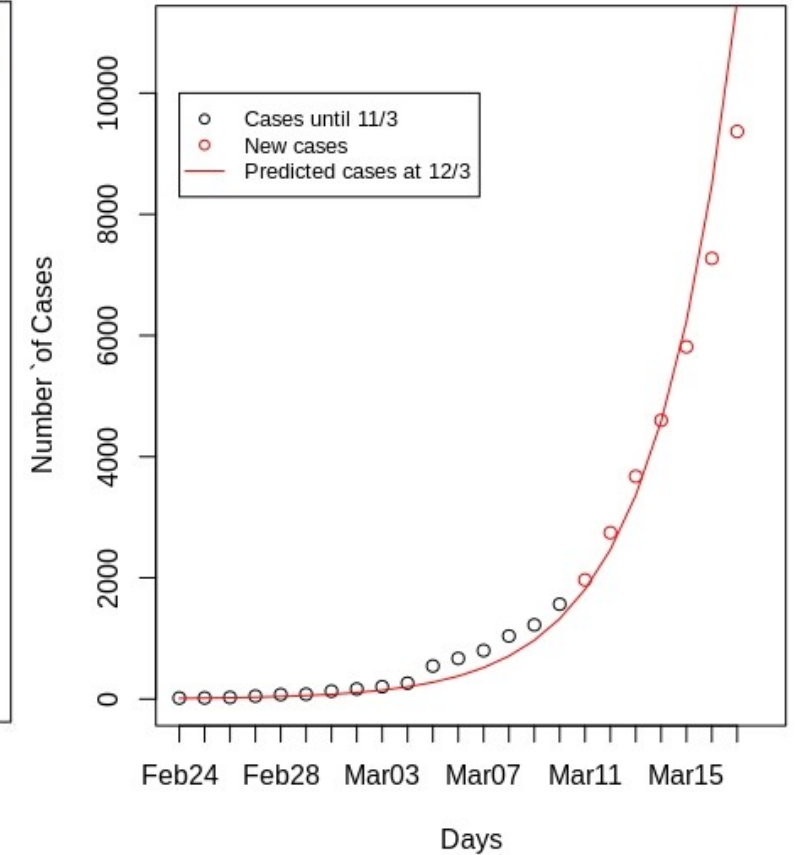
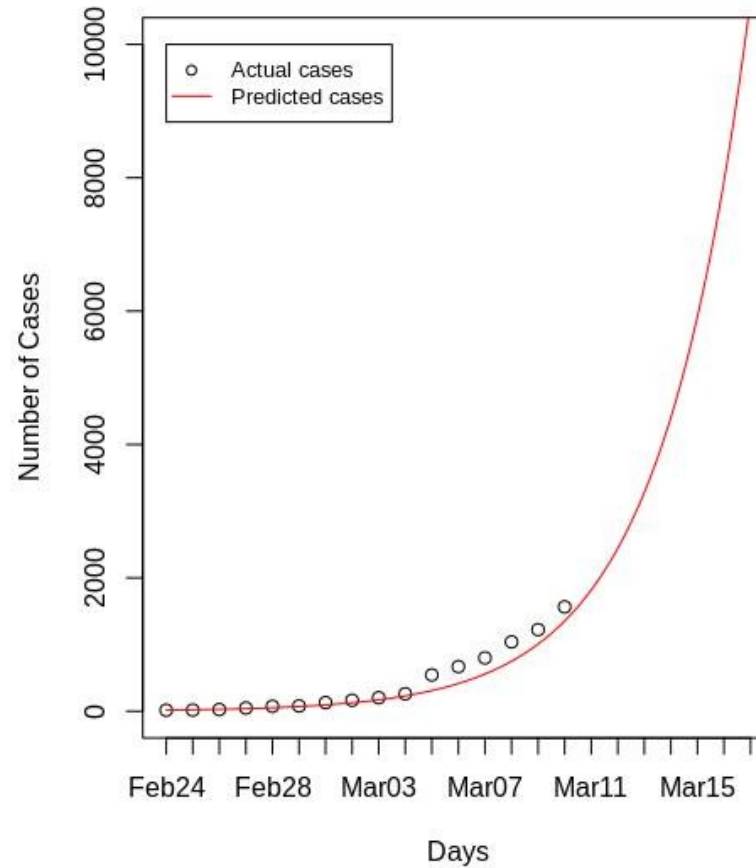
Linear models

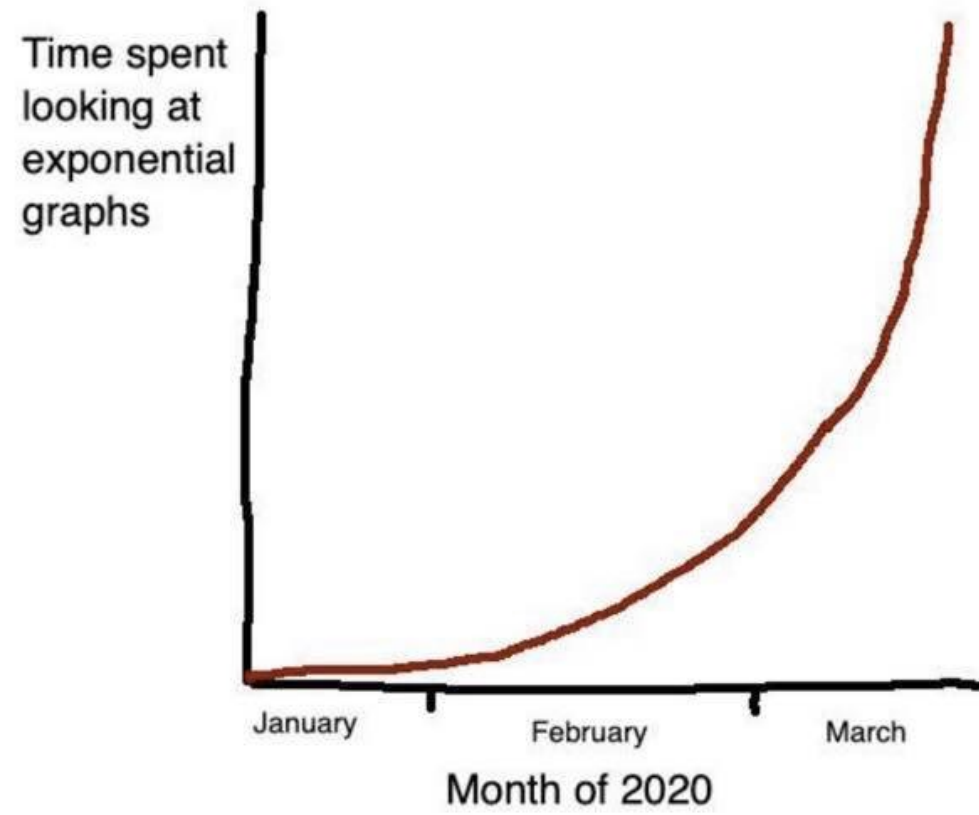
It is a linear model because the parameters combine additively; it is the model that needs to be linear, not necessarily the relationship(s)

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

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

Covid-19 cases in Germany

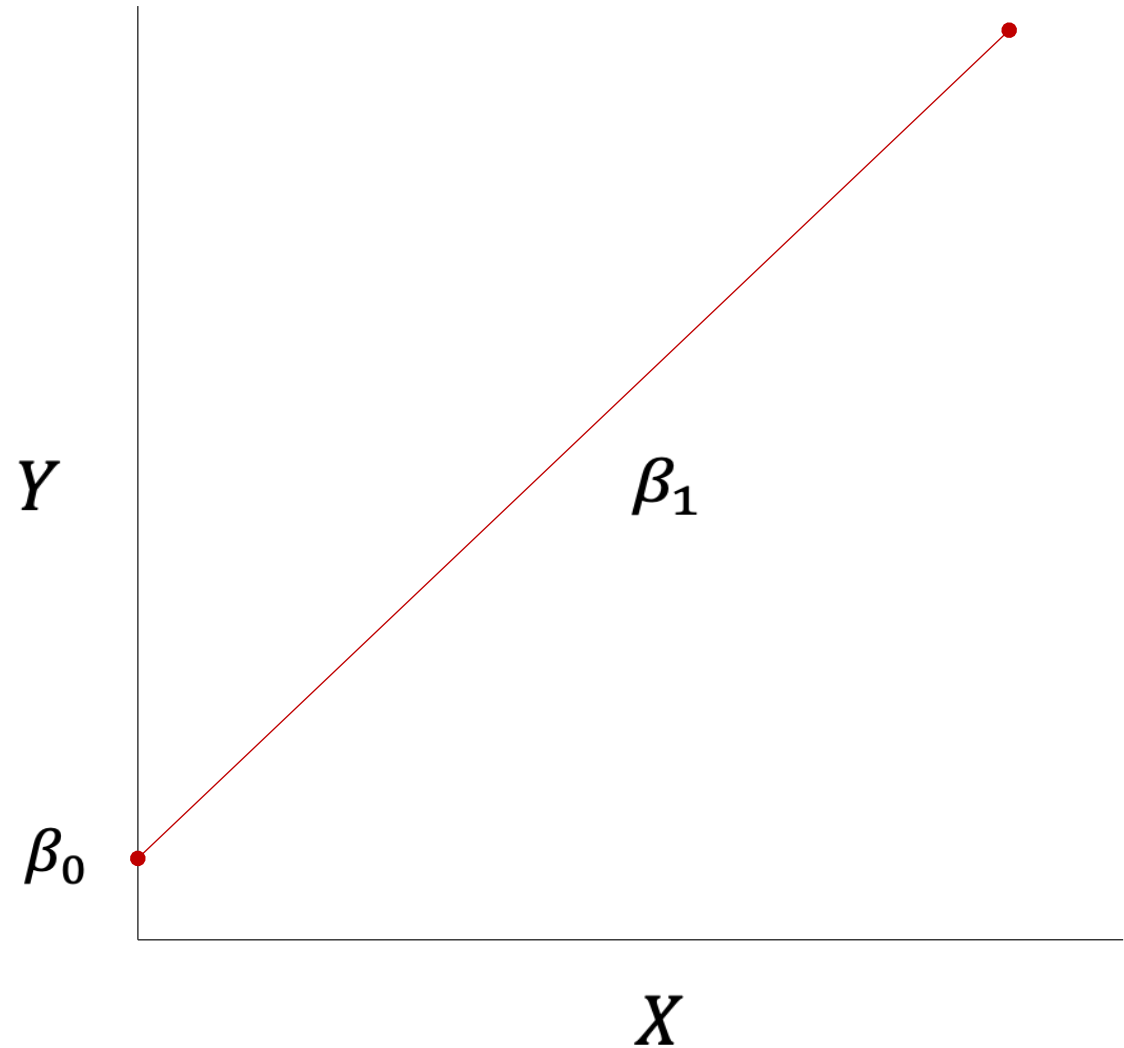




Linear models

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.

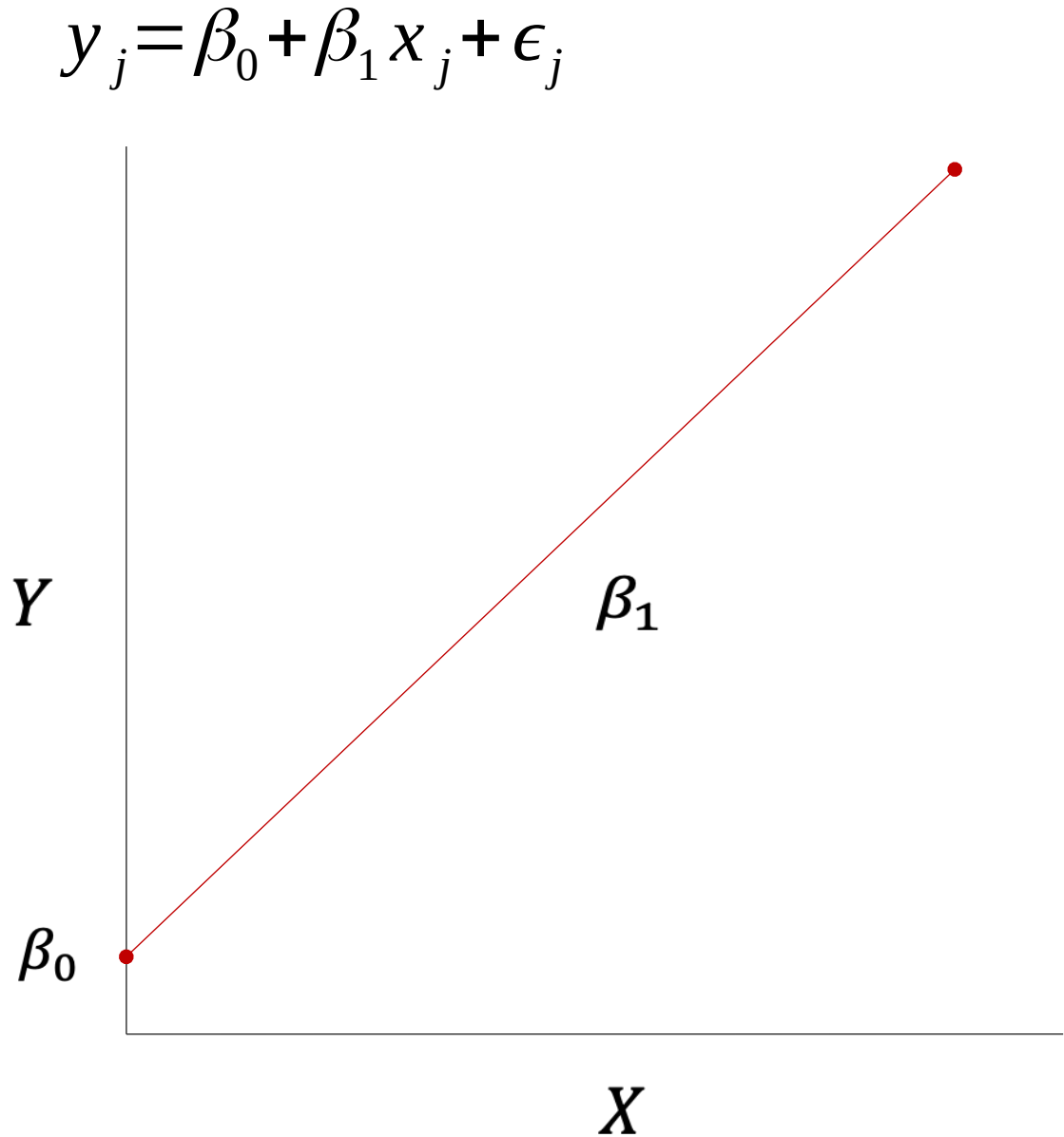
$$y_j = \beta_0 + \beta_1 x_j + \epsilon_j$$



Linear models

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.



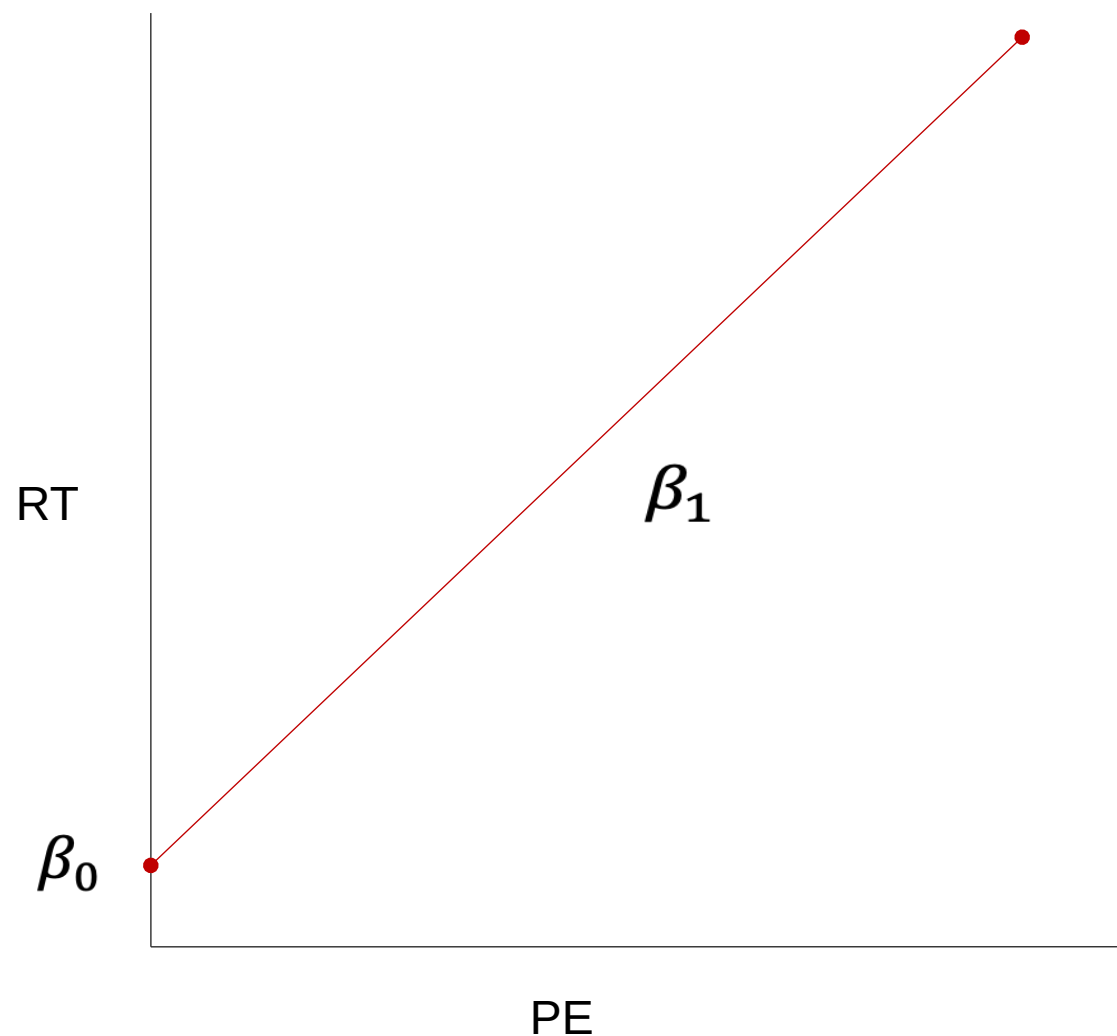
Linear models

What if we have variables that vary within-participants?



```
Lmmworkshop.Rmd  
```\{r simulate Data\}  
```
```

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

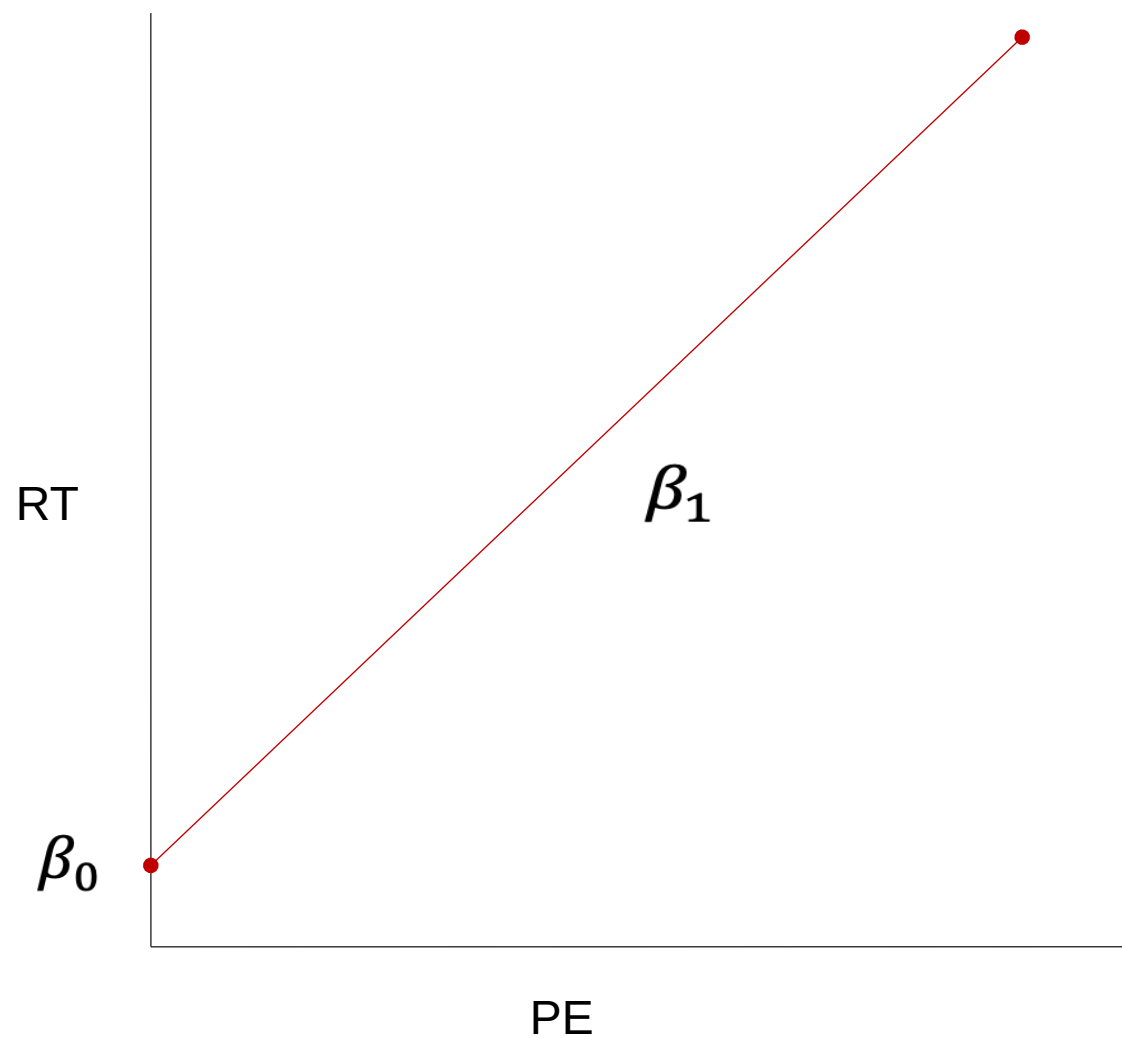


Linear models

What if we have variables that vary within-participants?

| subj_id | item_id | PE | RT |
|---------|---------|-------------|-----------|
| 1 | 1 | -0.38970130 | 8.683883 |
| 1 | 2 | 0.49204276 | 11.374542 |
| 1 | 3 | 0.22624263 | 10.577077 |
| 1 | 4 | 0.24938532 | 10.497849 |
| 1 | 5 | 0.12077736 | 9.852015 |
| 1 | 6 | -0.55368358 | 7.128669 |
| 1 | 7 | 0.32877114 | 11.021872 |
| 1 | 8 | -0.40407493 | 8.419886 |
| 1 | 9 | -0.22943074 | 9.080270 |
| 1 | 10 | 0.81836134 | 12.365893 |
| 1 | 11 | 0.64653442 | 11.805352 |
| 1 | 12 | 0.12154372 | 9.850657 |
| 1 | 13 | -0.43842637 | 7.905140 |
| 1 | 14 | -0.31403199 | 8.625142 |
| 1 | 15 | 0.81089257 | 12.933477 |
| 1 | 16 | 0.08442081 | 9.810517 |
| 1 | 17 | 0.33247623 | 10.117680 |
| 1 | 18 | -0.29181270 | 8.942813 |

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



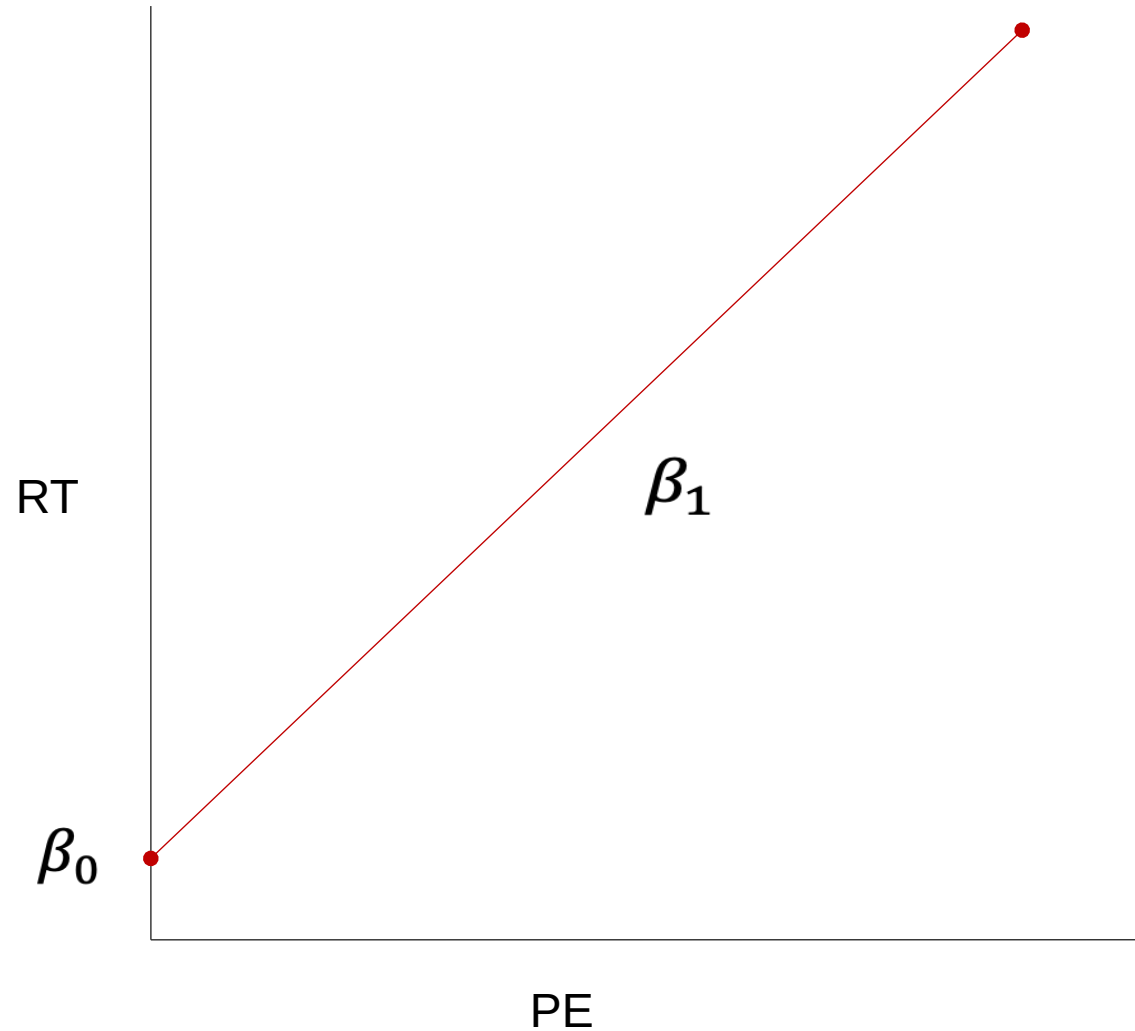
Linear models

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 independence, because participants provide more than one data point. Therefore, their data are correlated (there are clusters in the data).

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



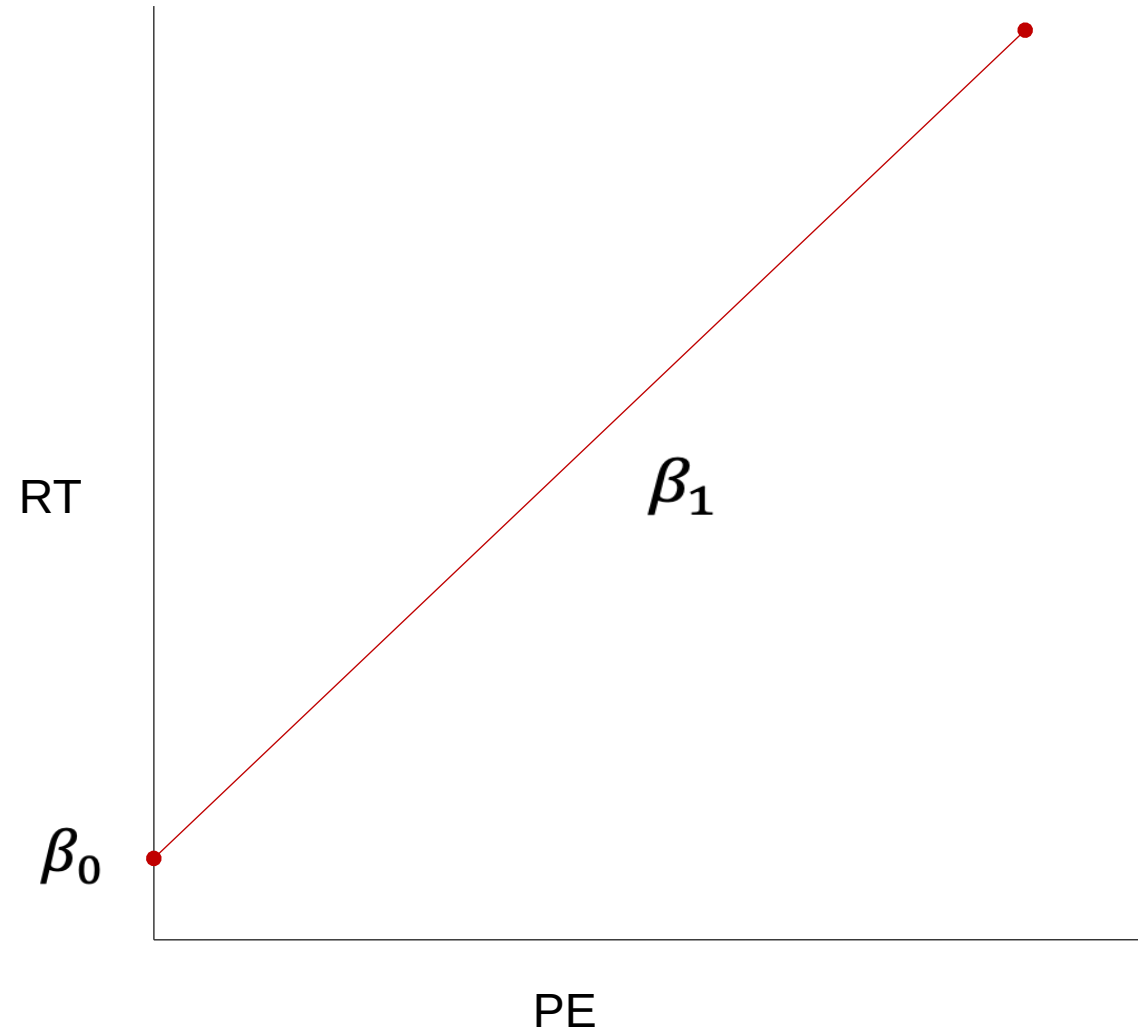
Linear models

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

```
Lmmworkshop.Rmd  
```\{r aggregate\}  
```
```

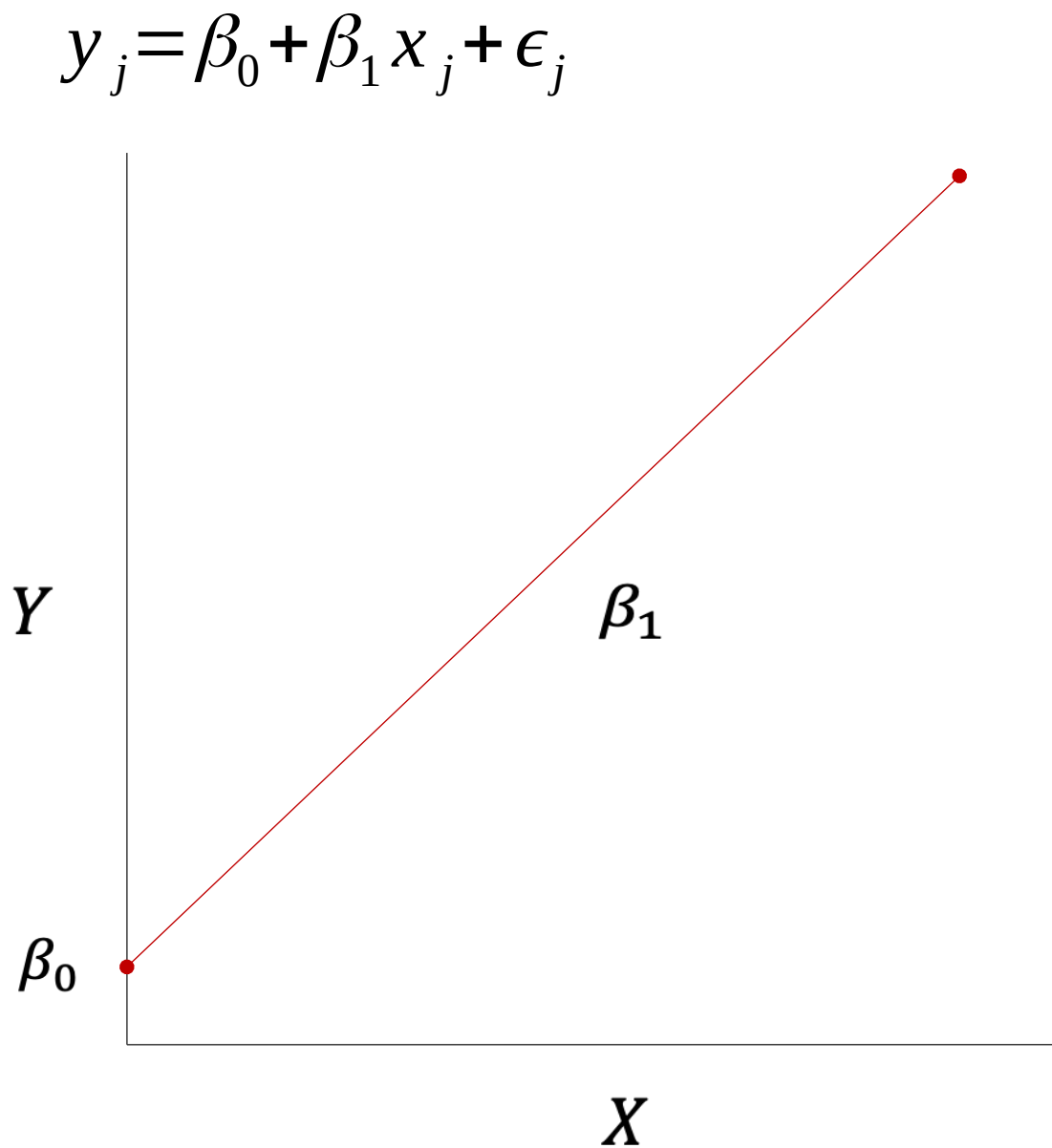


$$y_j = \beta_0 + \beta_1 x_j + \epsilon_j$$

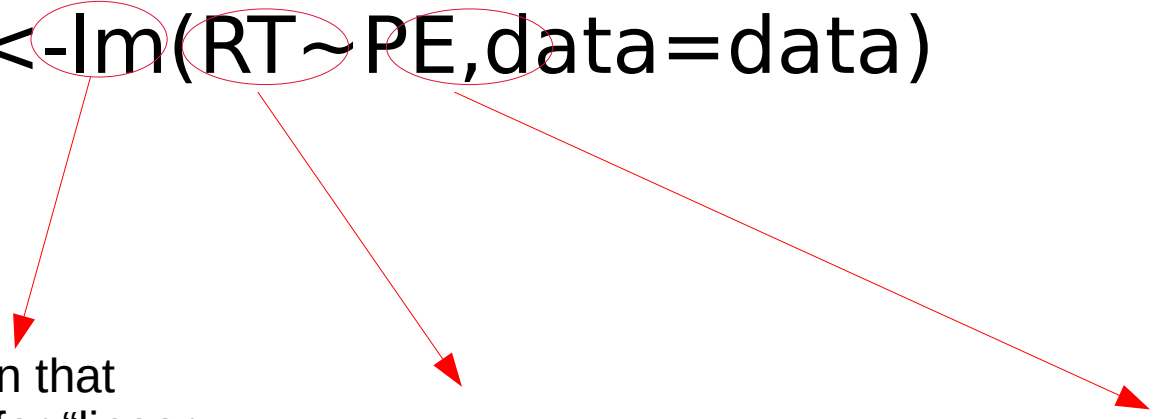


Linear models

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



LM<-lm(RT~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)
```

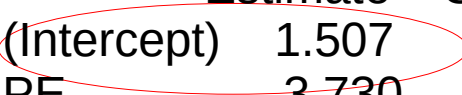
$$y_j = \beta_0 + \beta_1 x_j + \epsilon_j$$


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) | 1.507 | 7.024 | 0.215 | 0.832 |
| PE | -3.730 | 48.573 | -0.077 | 0.939 |



Residual standard error: 1.905 on 28 degrees of freedom

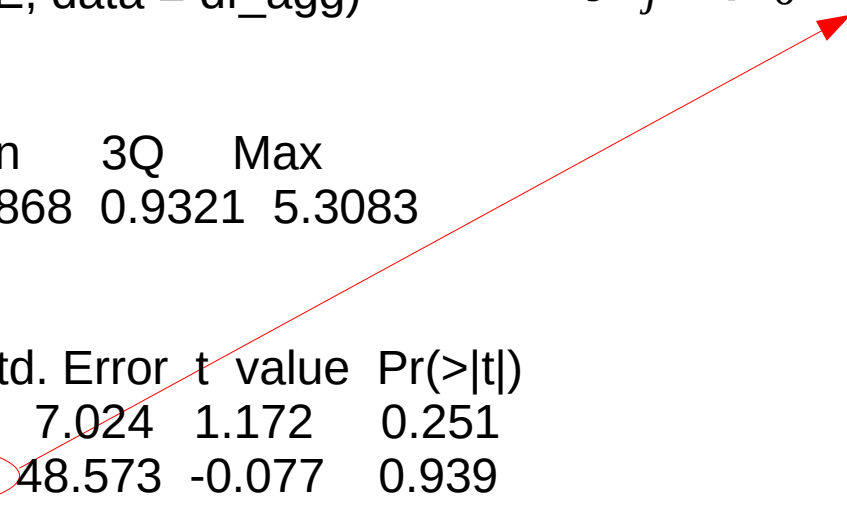
Multiple R-squared: 0.0002105, Adjusted R-squared: -0.0355

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 |

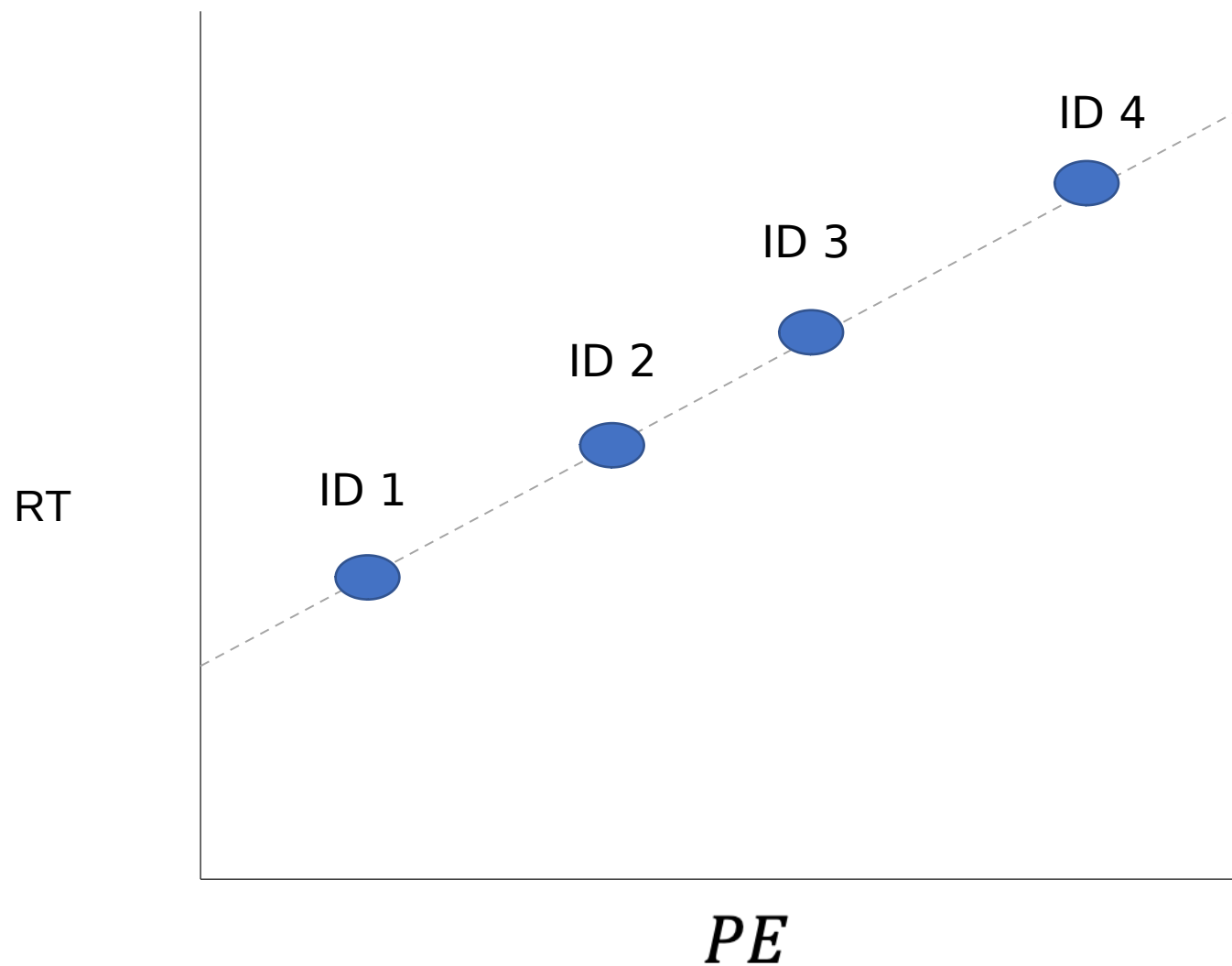
Coefficients:

| | Estimate | Std. Error | t value | Pr(> t) |
|-------------|----------|------------|---------|----------|
| (Intercept) | 1.507 | 7.024 | 0.215 | 0.832 |
| PE | -3.730 | 48.573 | -0.077 | 0.939 |

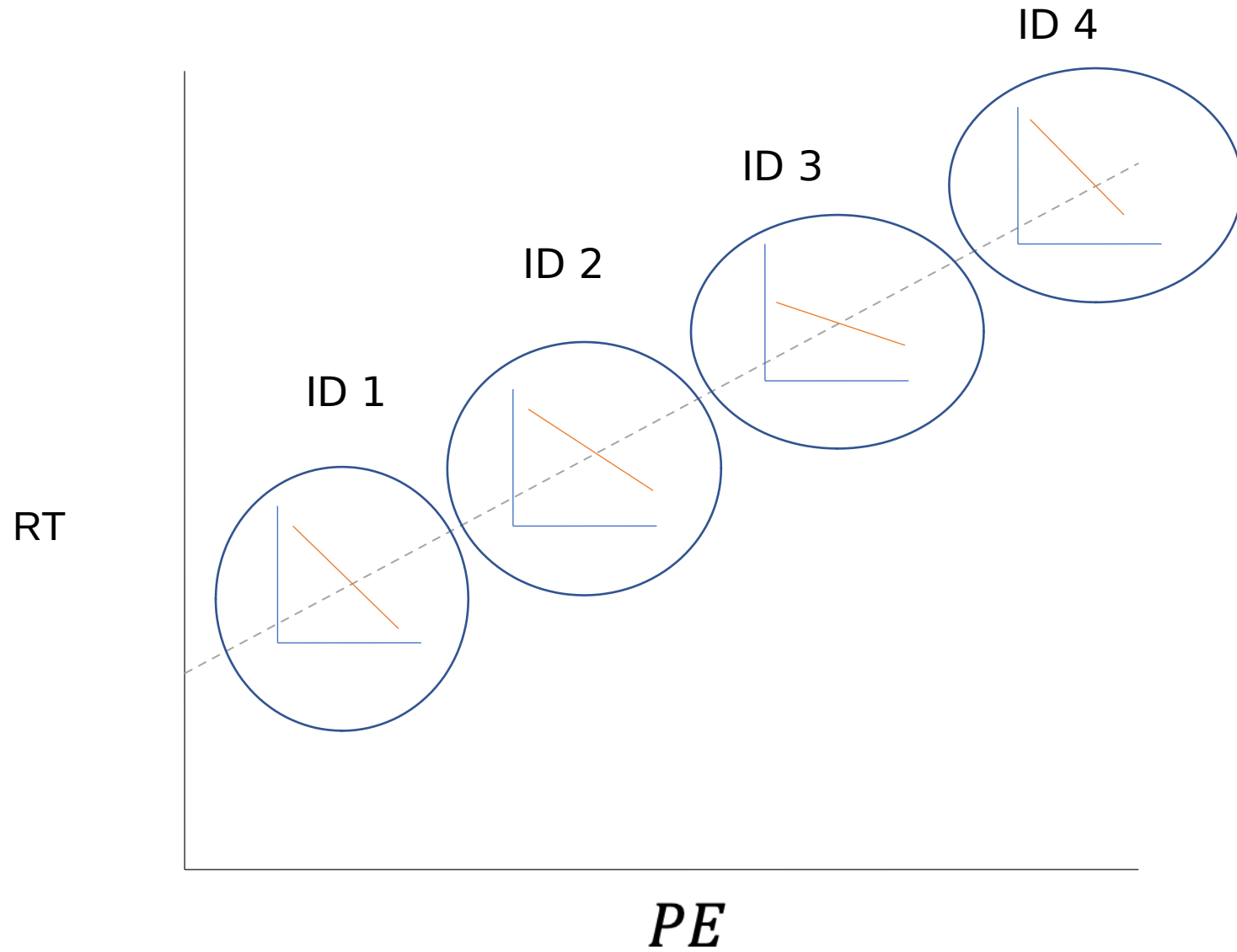
Residual standard error: 1.905 on 28 degrees of freedom
Multiple R-squared: 0.0002105, Adjusted R-squared: -0.0355
F-statistic: 0.005897 on 1 and 28 DF, p-value: 0.9393

Linear models

What is the problem with that?



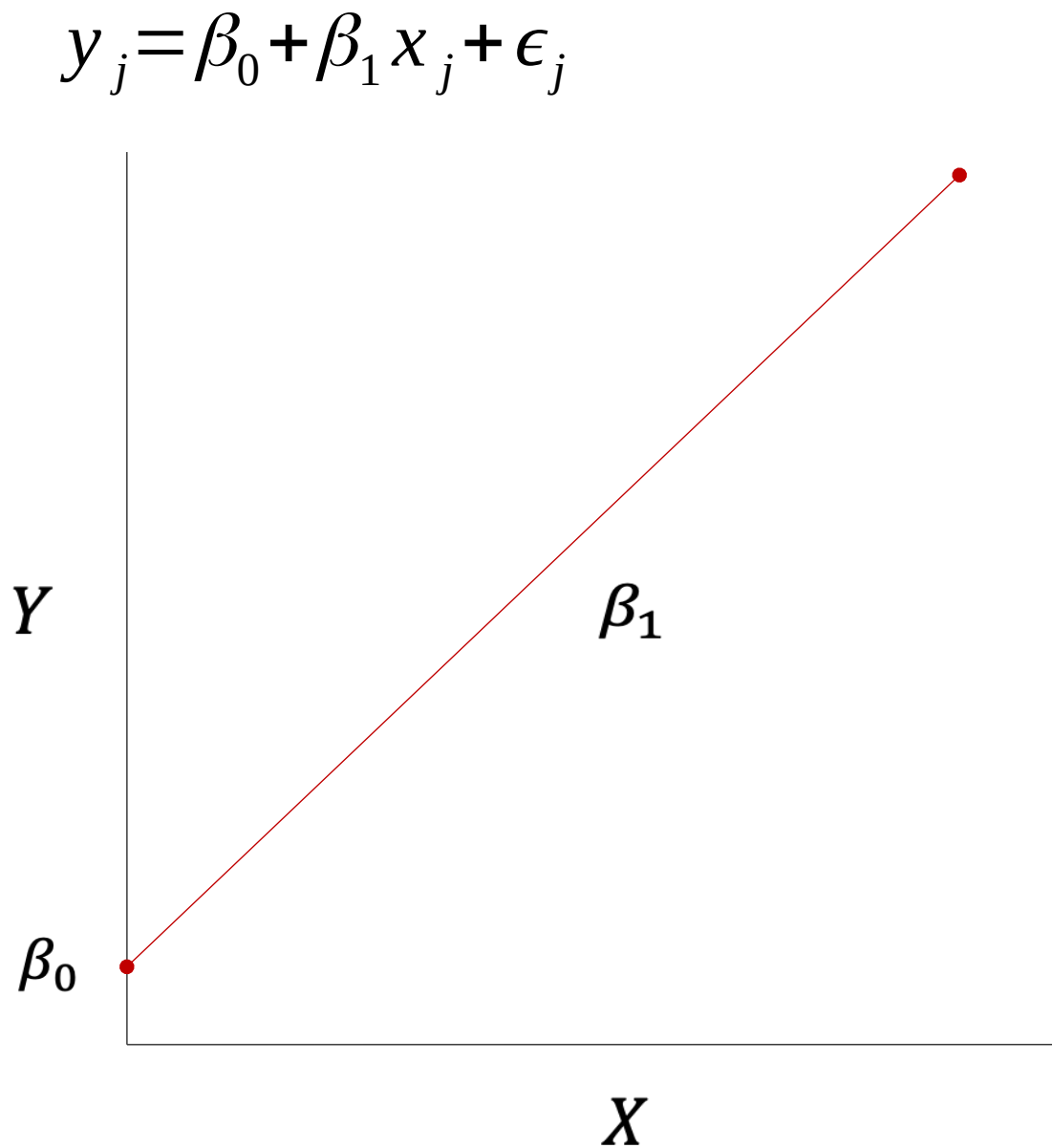
Between-participant
level:
Participants with
higher *PE* on average
have slower *RT*



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

Linear models

So what do we do?

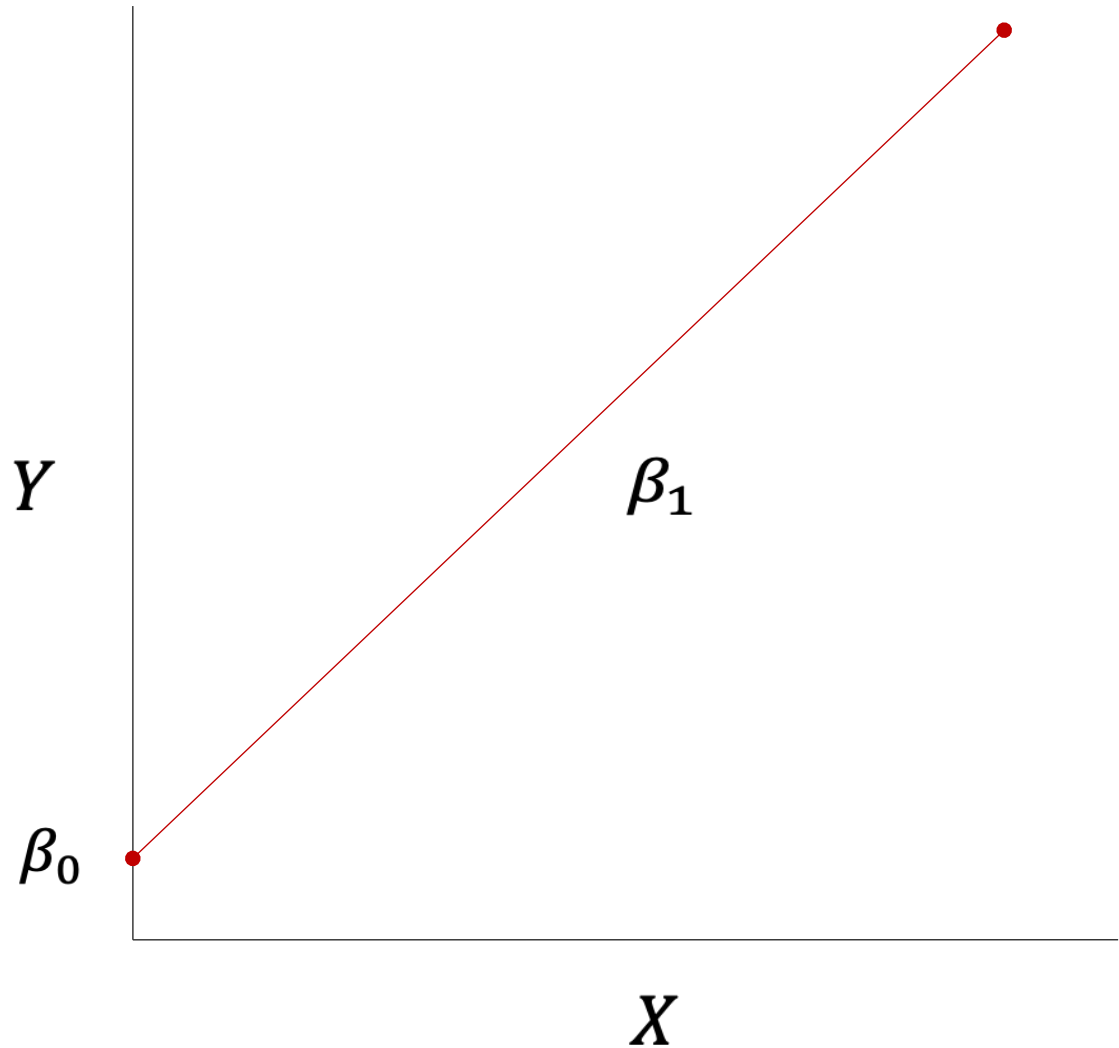


Linear mixed models

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 as being randomly selected from the larger population (Clark, 1973).

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

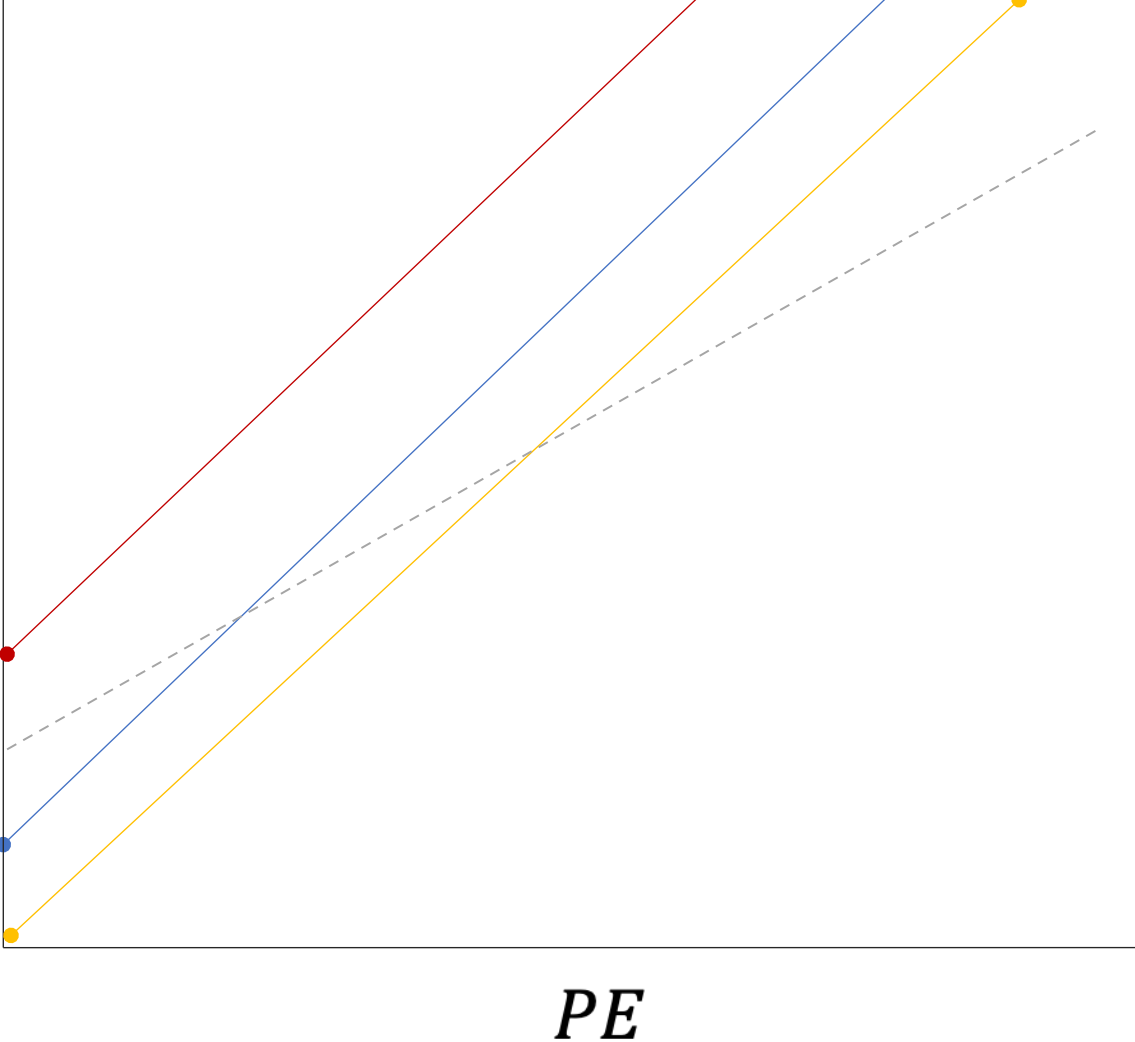


Random Intercepts

- Subject 1
- Subject 2
- Subject 3

(i.e., differences in “overall” RT)

RT

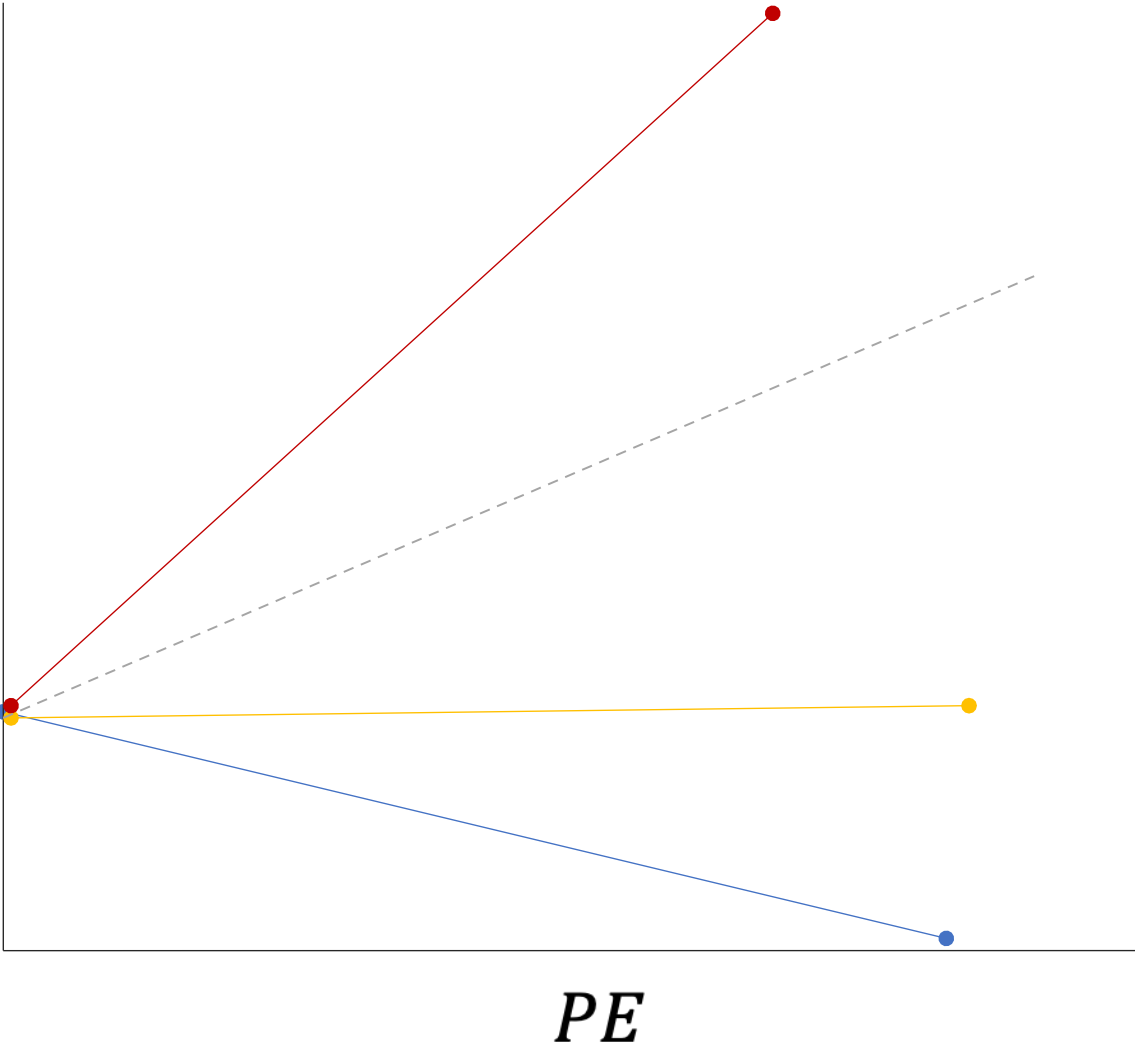


Random Slopes, fixed intercepts

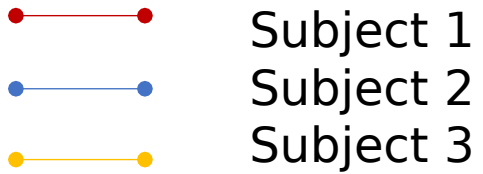
- Subject 1
- Subject 2
- Subject 3

RT

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

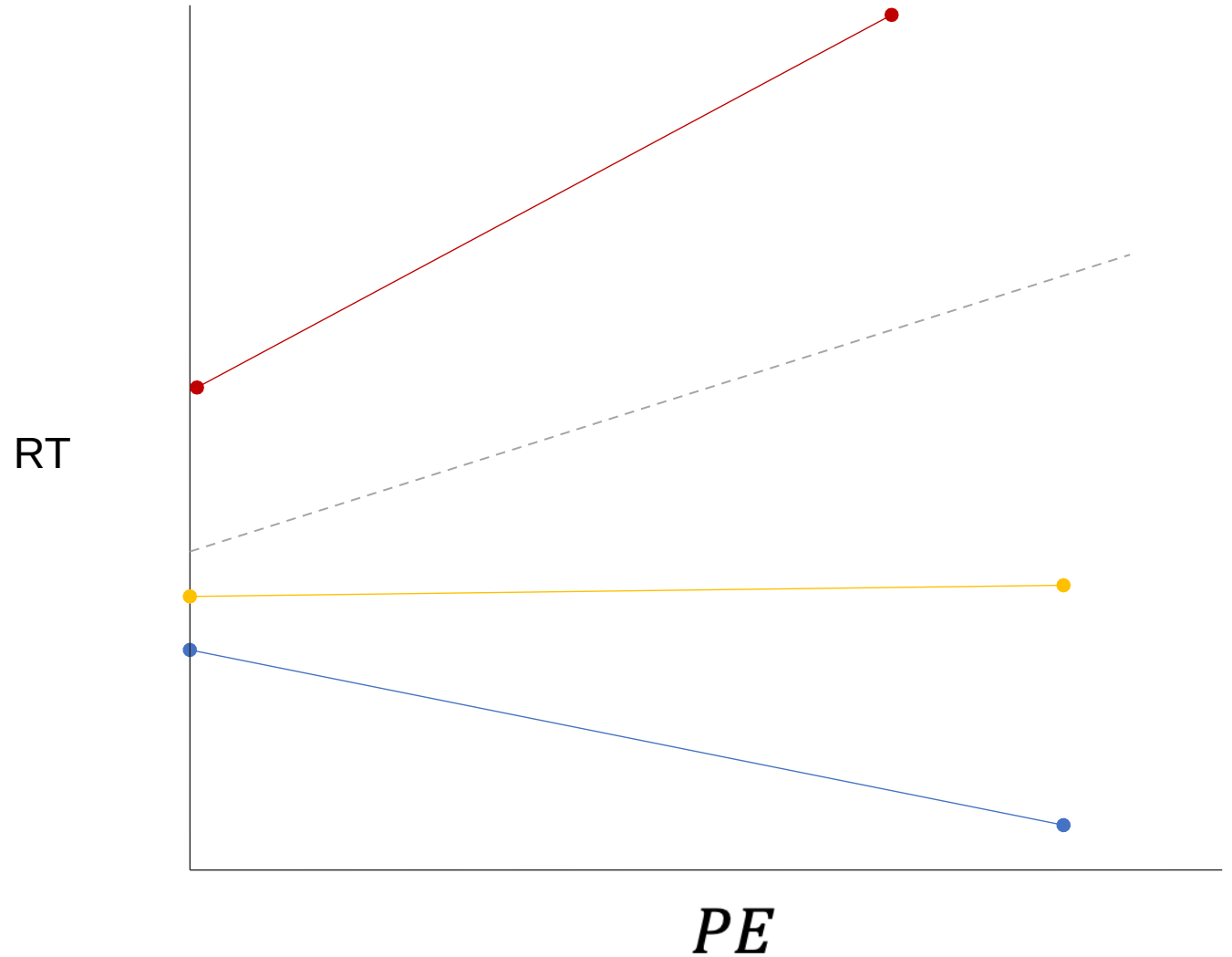


Random Intercepts and Slopes

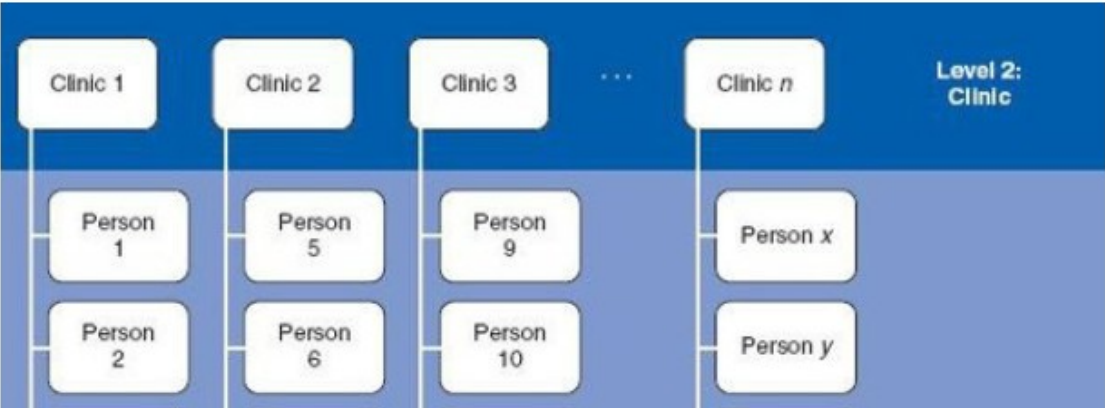


(i.e., differences in the size of the effect across subjects)

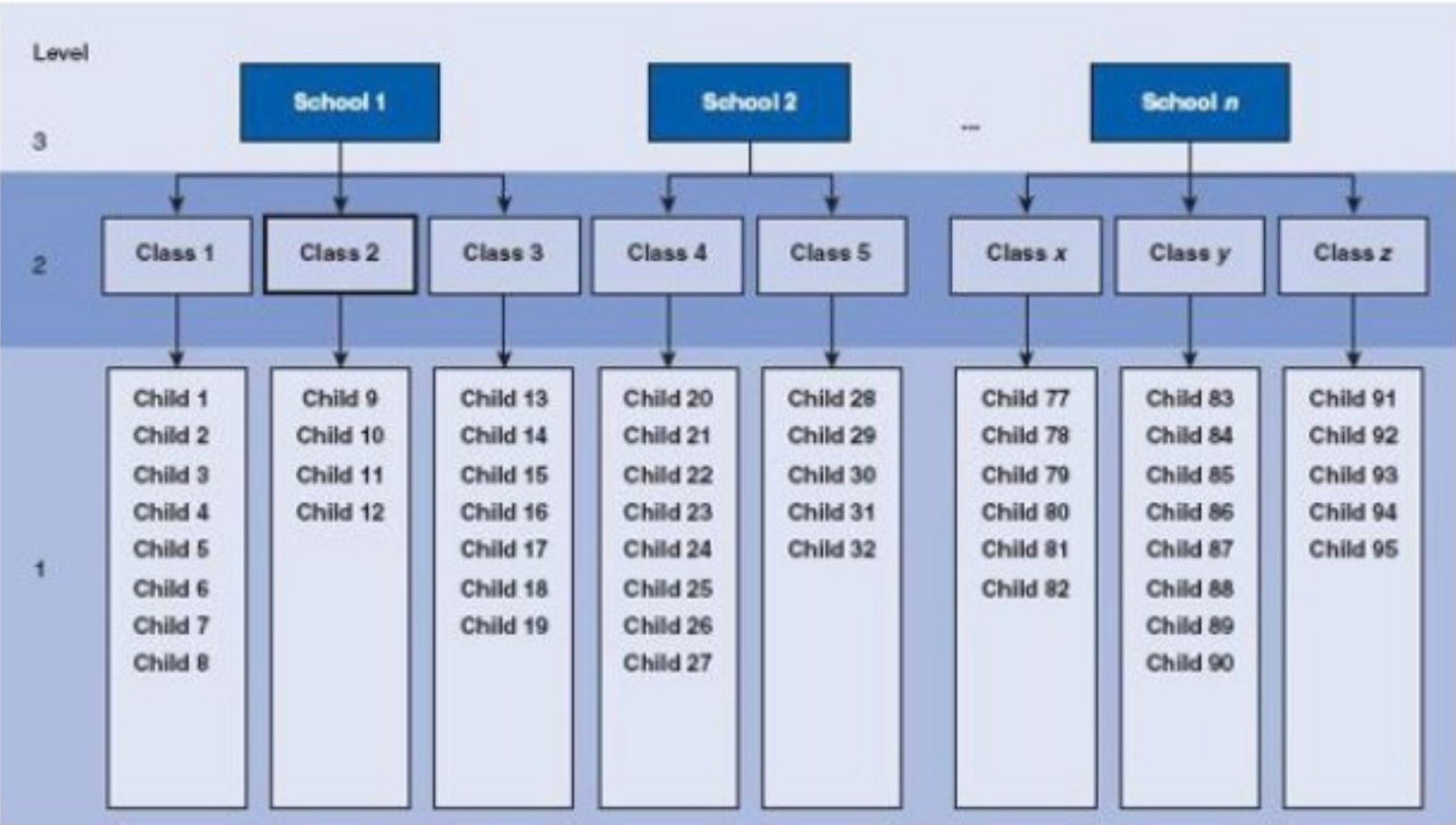
Including random intercepts and slopes allows to control for random variations among items (sampling units)



Linear mixed models
Multilevel models
Hierarchical linear mod



Nested data



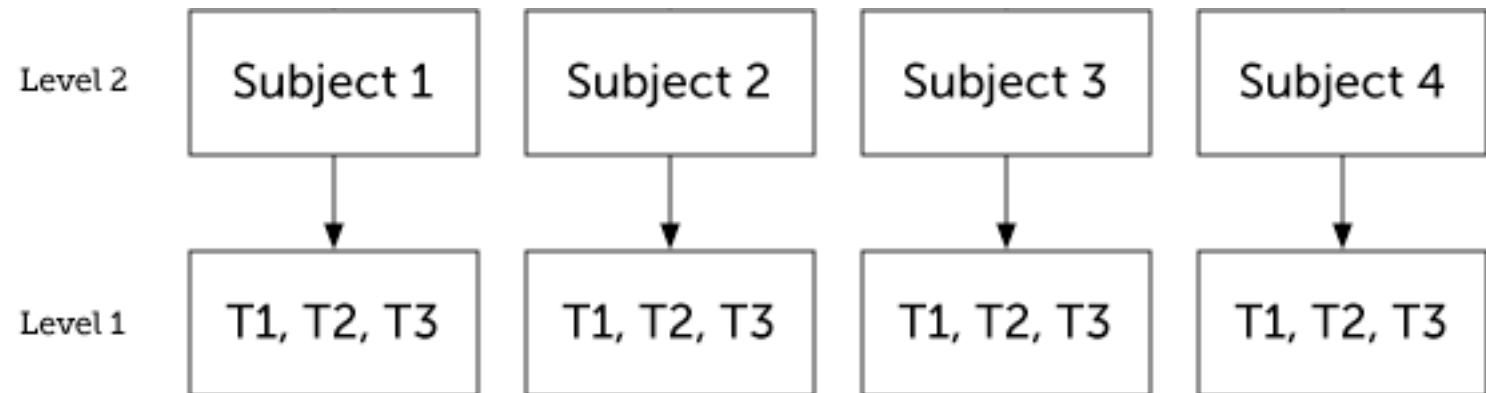
Linear mixed models

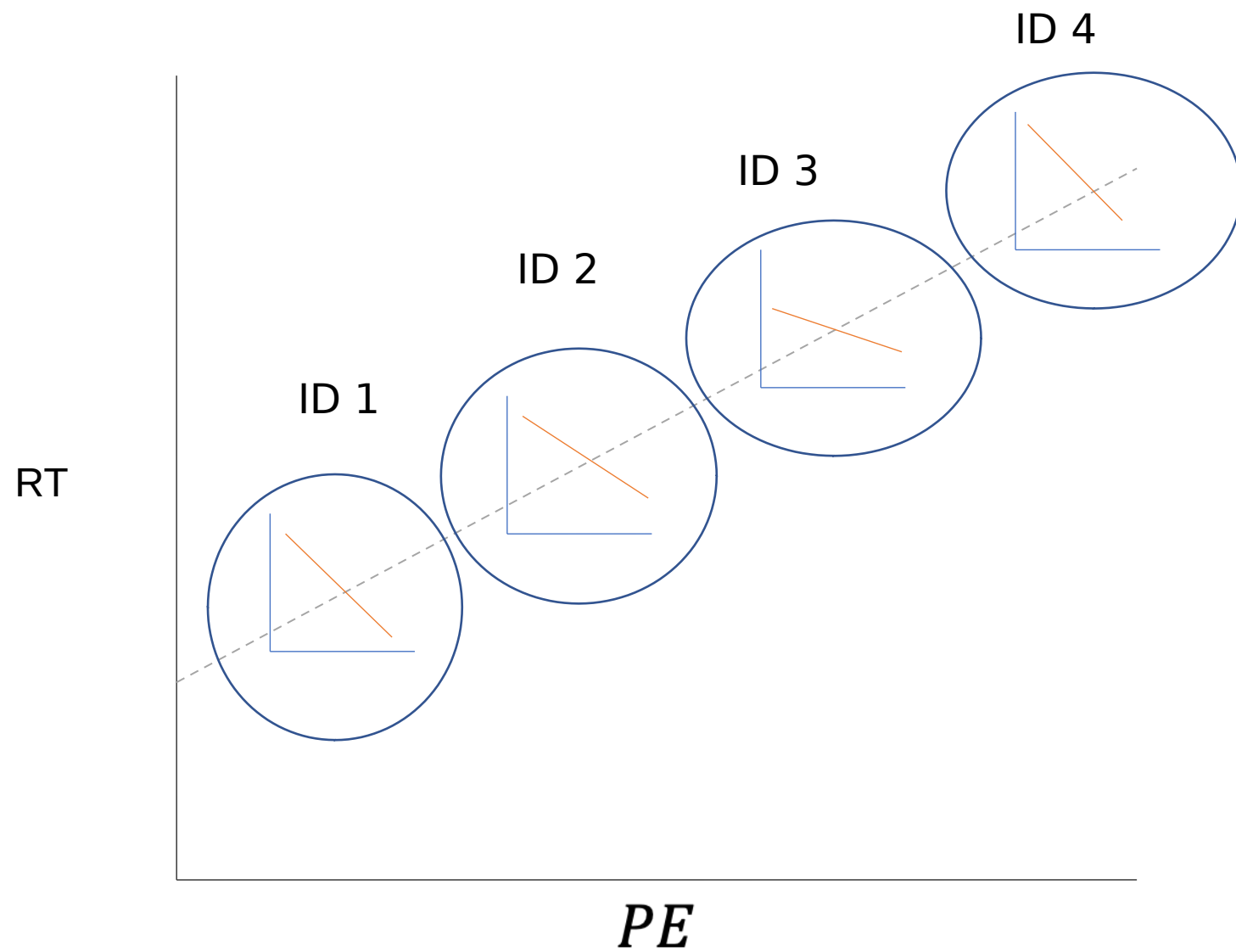
Multilevel models

Hierarchical linear models

Nested data

Taking into account
within- and between-
person variability



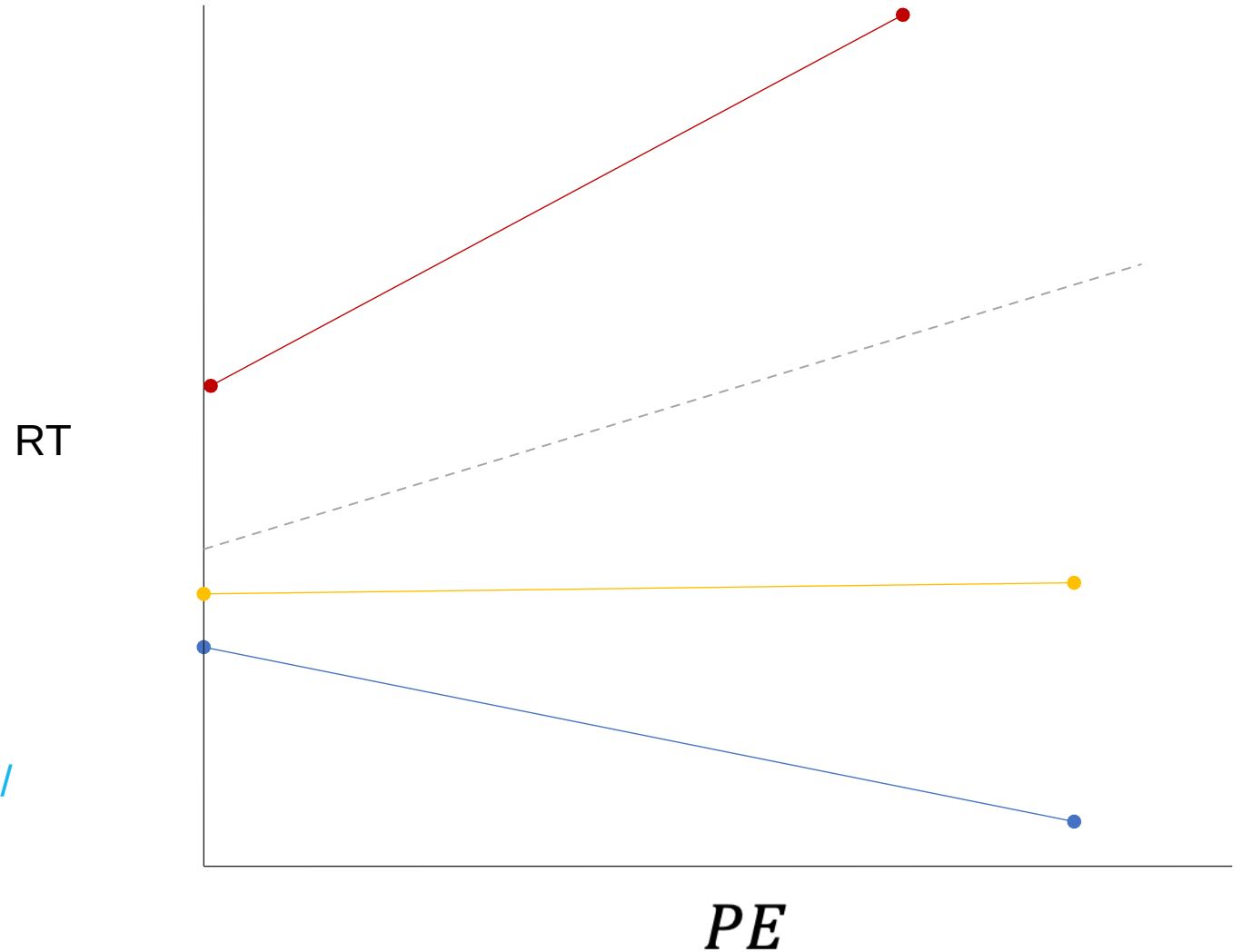
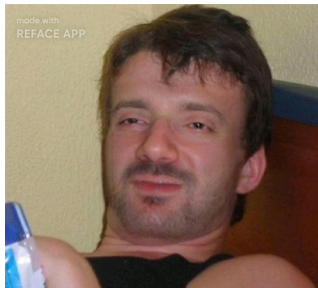


“The reason for favoring the within-subject level is that omitted and confounding variables are less likely to be a problem when analyses focus on how and why people change over time than on how people differ from one another”. (Bolger & Laurenceau, *Intensive longitudinal methods: An introduction to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



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

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

The dependent variable Y for subject j for the specific observation i is given by an subject-specific intercept β_{0j} , a subject specific slope β_{1j} , and a within-subject error term ϵ_{ij} .

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

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

Level 2
(population level)

u_{0j} is a random effect because it varies from cluster to cluster

Subject 1

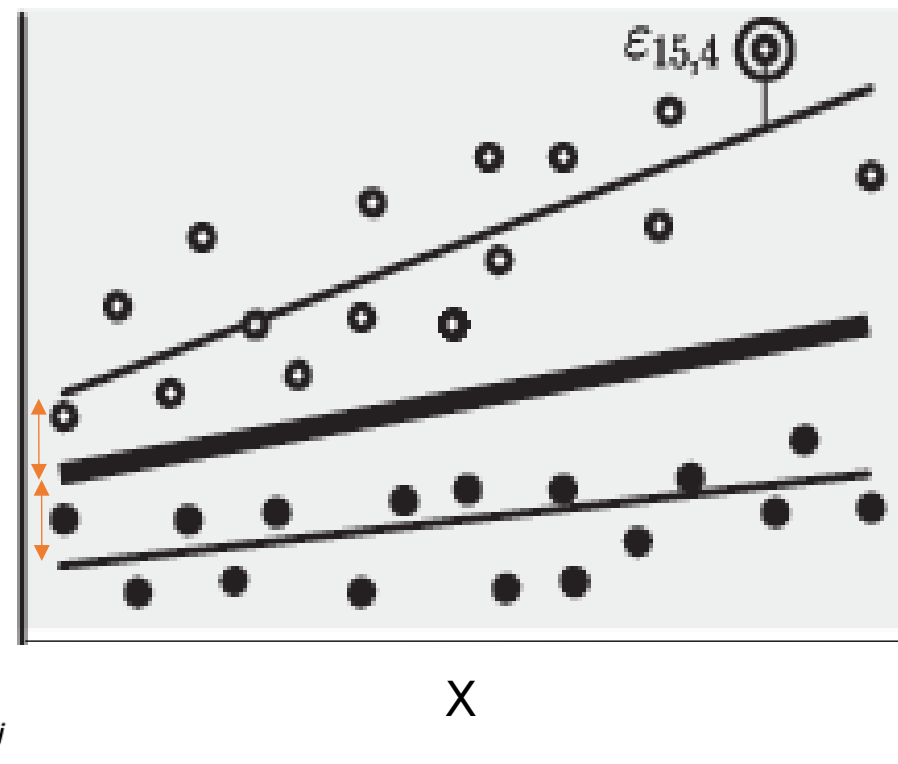
$$Y_{i1} = (\gamma_{00} + u_{01})$$

Typical person

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

Subject 2

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



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

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

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

Level 2
(Population level)

Subject 1

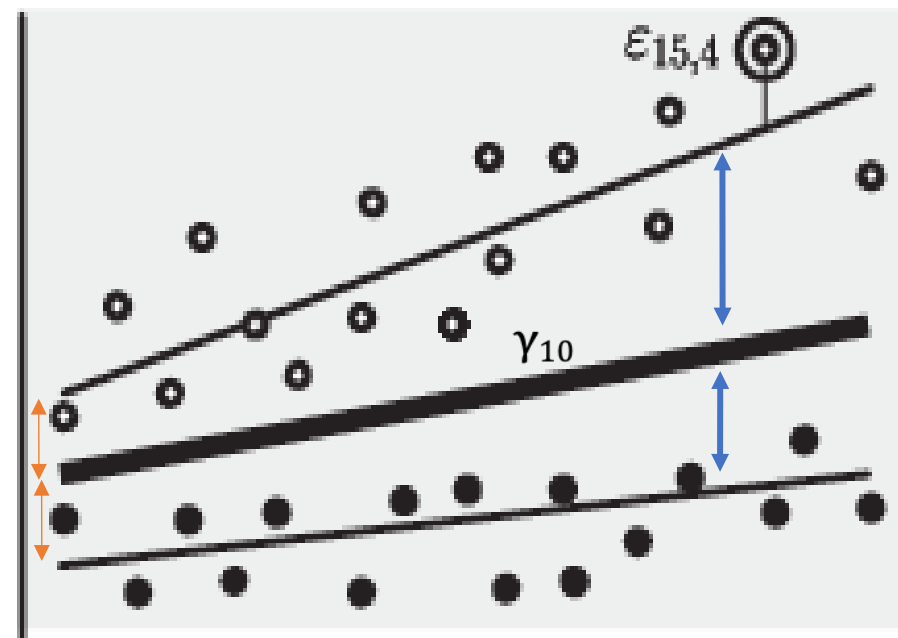
$$Y_{i1} = (\gamma_{00} + u_{01}) + (\gamma_{10} + u_{11}) X_{i1} + \epsilon_{i1}$$

Typical person

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

Subject 2

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



X

u_{0j}

u_{1j}

$Y_{ij} = \beta_{0j} + \beta_{1j} X_{ij} + \epsilon_{ij}$
 is a within-subject residual term representing the difference, at a given time point between the predicted Y for a given subject and the actual value.

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

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

Level 2 (between-subjects)

Subject 1

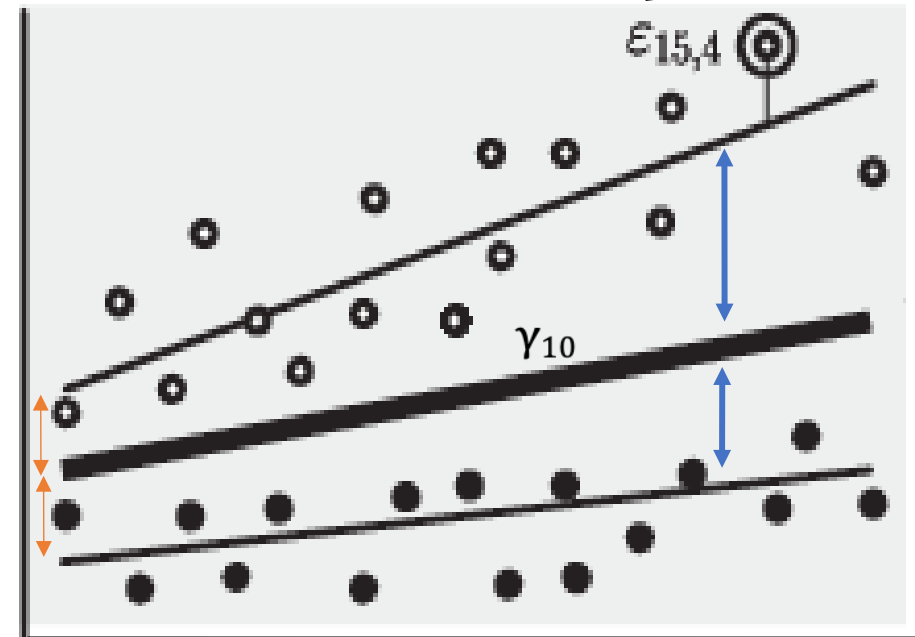
$$Y_{i1} = (\gamma_{00} + u_{01}) + (\gamma_{10} + u_{11}) X_{i1} + \epsilon_{i1}$$

Typical person

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

Subject 2

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



X

u_{0j}

u_{1j}

$Y_{ij} = \beta_{0j} + \beta_{1j} X_{ij} + \epsilon_{ij}$
 is a within-subject residual term representing the difference, at a given time point between the predicted Y for a given subject and the actual value.

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

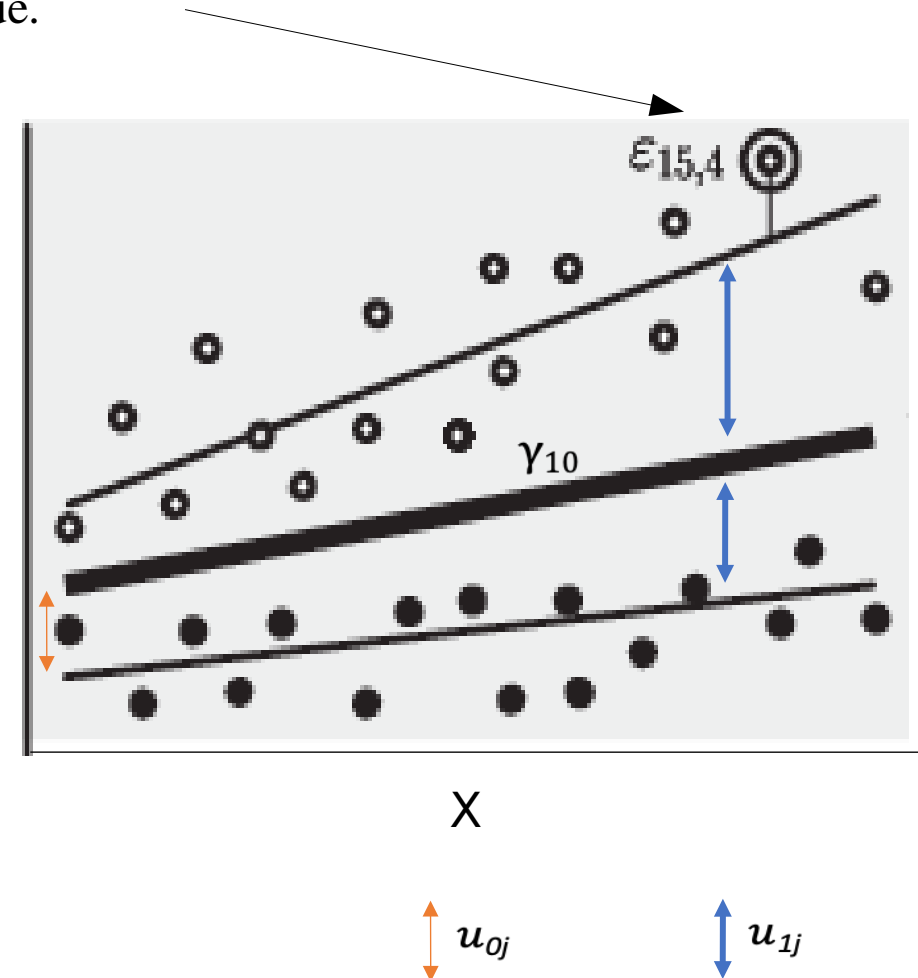
$$\beta_{1j} = \gamma_{10} + u_{1j}$$

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

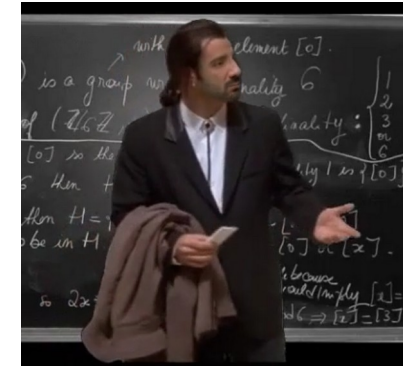
$$Y_{ij} = \beta_{0j} + \beta_{1j} PE_{ij} + \epsilon_{ij} \quad \text{Level 1}$$

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

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

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

If we set the us to 0, we have a fixed effects model



Model formulation

$$Y_{ij} = \beta_{0j} + \beta_{1j} X_{ij} + \epsilon_{ij} \quad \text{Level 1}$$

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

Level 2

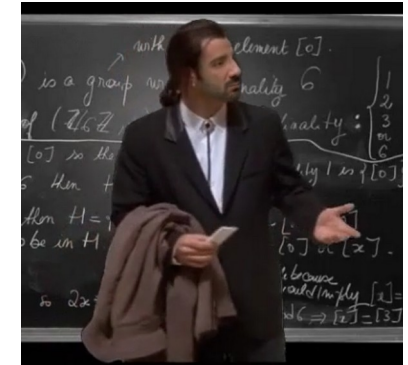
$$\beta_{1j} = \gamma_{10} + u_{1j}$$

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

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 algorithms.

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 Likelihood? 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

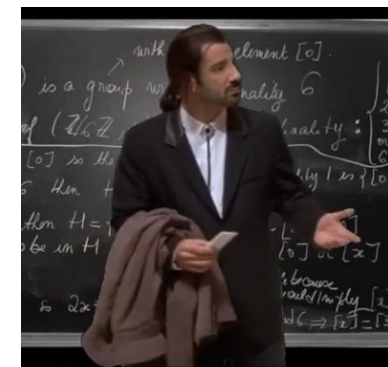
τ^2 = Population variance between clusters

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 than 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)

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

τ_{0j}^2

σ_{1j}^2

Lmmworkshop.Rmd
`\\{r lmm rand int}`



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

$$\hat{\rho} = \frac{3.49}{3.49 + 4.03} = 0.46$$

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:

- Akaike 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.

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



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

Data: df

Models:

mod_unc: RT ~ 1

mixmod_unc: RT ~ 1 + (1 | subj_id)

| | npars | 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

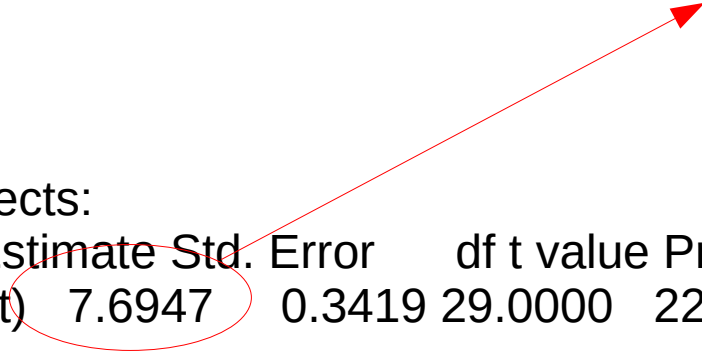
If models are not fitted with ML, they are refitted with ML, because it makes them comparable.

The chi square difference test is computed on deviance statistics, which is $-2 \cdot LL$, 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 overfitting 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

τ_{00}^2
Variation In RT across subjects

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

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

τ_{10}^2

Variation in the effect on PE on RT across subjects

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

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 between random intercepts and random slopes



y_{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: subtracting the grand mean from the person mean

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

Lmmworkshop.Rmd
``{r centering}
``



Pseudo R²

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)

R²_{pseudo:within}

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

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

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

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 R²

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)

R²_{pseudo:between}

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

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

Intercept variance mixmod_unc Intercept variance ModPE

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

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 R²

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)

R²pseudo:bw (between and within)

- How much variance of criterion can be explained by predictor

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

- 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

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

Level 2

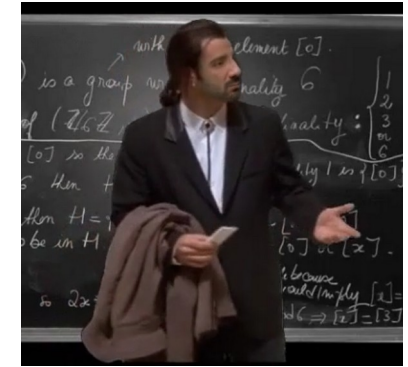
$$\beta_{1j} = \gamma_{10} + u_{1j}$$

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

Variance-covariance matrix - unstructured

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

Residuals



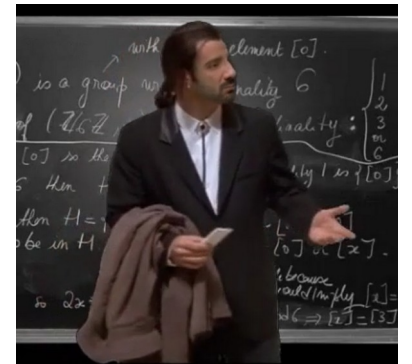
Variance-covariance structures

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

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 independent, so their covariance is set at 0



Model Selection

You can start with 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 long as you add a clear statement of the criteria 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 $\alpha = 0.05$ 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, $\alpha = 0.2$

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}  
``
```



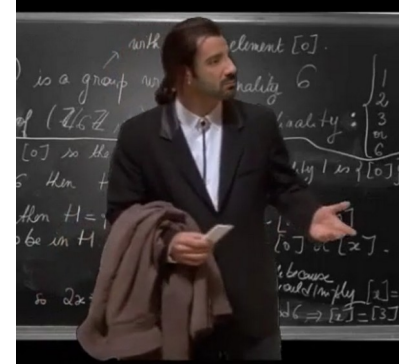
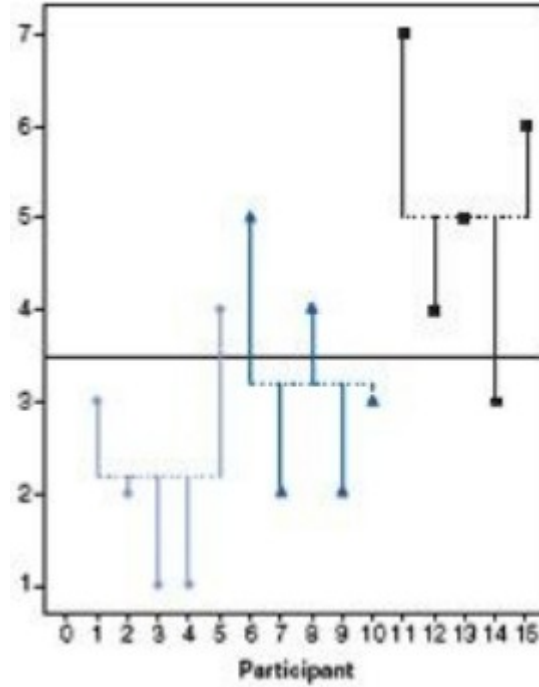
Mod0 = ?

Mod1 = ?

Mod2 = ?

categorical predictor – ANOVA

$$Y_i = \beta_0 + \beta_1 PE_{low} + \beta_2 PE_{high} + \epsilon_i$$

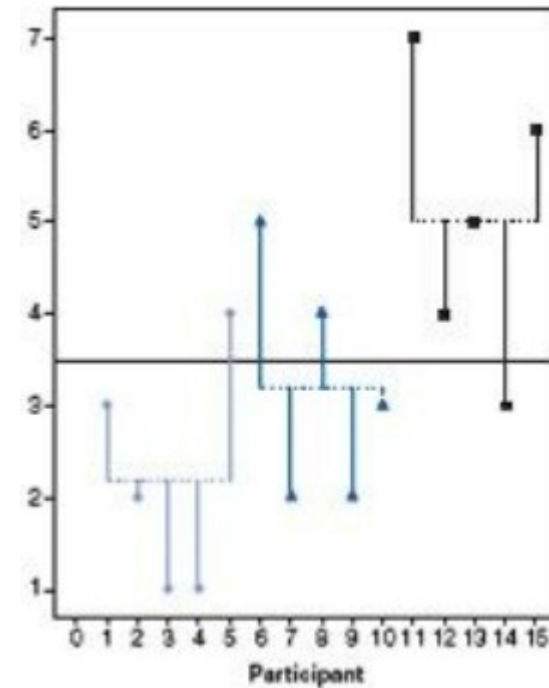


categorical predictor – ANOVA

$$Y_i = \beta_0 + \beta_1 PE_{low} + \beta_2 PE_{high} + \epsilon_i$$

Lmmworkshop.Rmd

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



**categorical predictor – between participants**

$$Y_i = \beta_0 + \beta_1 PE_{low} + \beta_2 PE_{high} + \epsilon_i$$

Coefficients:

	Estimate	td. Error	t value	Pr(> t )
(Intercept)	7.8886	0.5690	13.864	8.52e-14 ***
PEbwHighPE	0.5368	0.8047	0.667	0.510
PEbwLowPE	-1.1184	0.8047	-1.390	0.176

PEbwHighPE = PEHigh - PEMedium  
PEbwLowPE = PELow - PEMedium

$$\beta_0 = 7.89$$

$$\beta_1 = 7.89 - 1.12 = 6.77$$

$$\beta_2 = 7.89 + 0.54 = 8.43$$



categorical predictor – between participants

$$Y_i = \beta_0 + \beta_1 PE_{low} + \beta_2 PE_{high} + \epsilon_i$$

We can also set meaningful contrasts

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



Is it the same if we transform categorical into continuous?

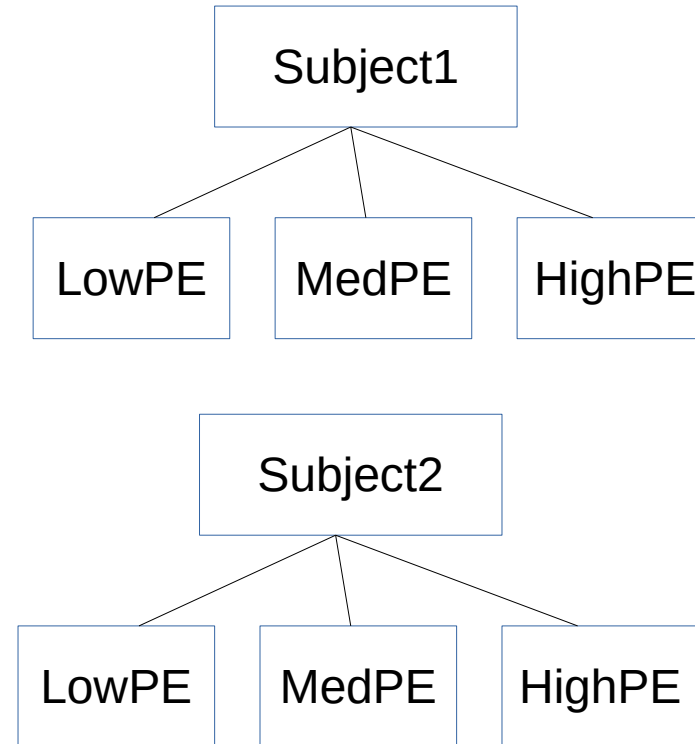
**Continuous variable:** if we treat a variable as continuous we are willing to treat the differences between values as quantitative. Thus, it is meaningful to talk about the effect of “ a one-point increase” in the values of X.

**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.

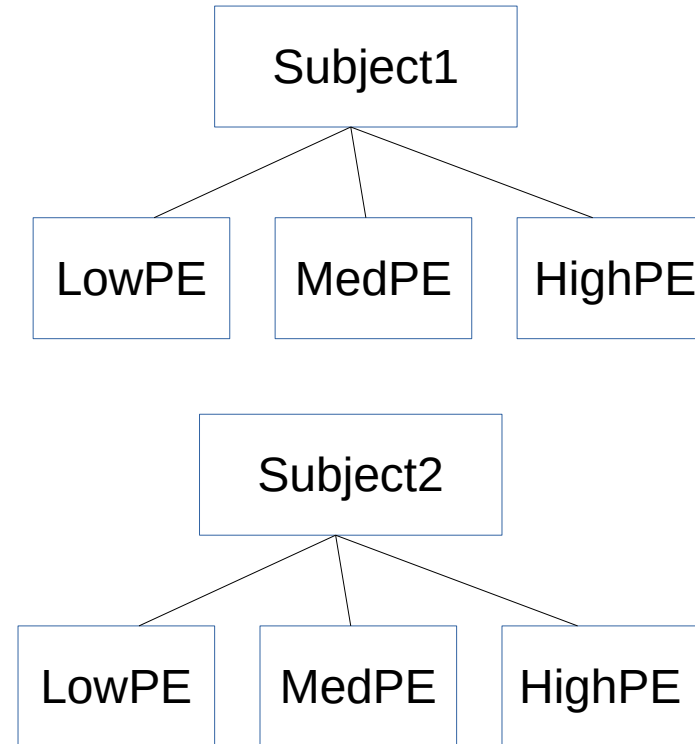
```
Lmmworkshop.Rmd
```${r continuous predictor between}``  
```
```



**categorical predictor – within participants**



## categorical predictor – within participants



Lmmworkshop.Rmd

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



categorical predictor – within participants, mixed effects

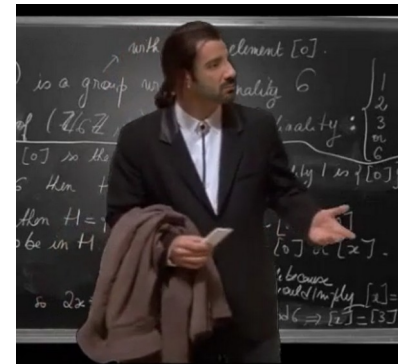
$$Y_{ij} = \beta_{0j} + \beta_{1j} PElow_i + \beta_{2j} PEhigh_i + \epsilon_{ij}$$

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

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

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

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



## categorical predictor – within participants, mixed effects

Lmmworkshop.Rmd

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



Random effects:

| Groups  | Name        | Variance | Std.Dev. | Corr       |
|---------|-------------|----------|----------|------------|
| subj_id | (Intercept) | 4.236    | 2.058    |            |
|         | PElevel2    | 3.389    | 1.841    | -0.86      |
|         | PElevel3    | 22.146   | 4.706    | -0.81 0.99 |

How much participants in the Pelevel1 condition (reference) vary in their average RT.

## categorical predictor – within participants, mixed effects

Lmmworkshop.Rmd

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



Random effects:

| Groups  | Name        | Variance | Std.Dev. | Corr       |
|---------|-------------|----------|----------|------------|
| subj_id | (Intercept) | 4.236    | 2.058    |            |
|         | PElevel2    | 3.389    | 1.841    | -0.86      |
|         | PElevel3    | 22.146   | 4.706    | -0.81 0.99 |

How much the difference in RT  
between PElevel2 and PElevel1 varied  
across participants

## categorical predictor – within participants, mixed effects

Lmmworkshop.Rmd

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



Random effects:

| Groups  | Name        | Variance | Std.Dev. | Corr       |
|---------|-------------|----------|----------|------------|
| subj_id | (Intercept) | 4.236    | 2.058    |            |
|         | PElevel2    | 3.389    | 1.841    | -0.86      |
|         | PElevel3    | 22.146   | 4.706    | -0.81 0.99 |

How much the difference in RT  
between PElevel3 and PElevel1 varied  
across participants



## **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 GLMMs
- Cutting-edge

Lmmworkshop.Rmd

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



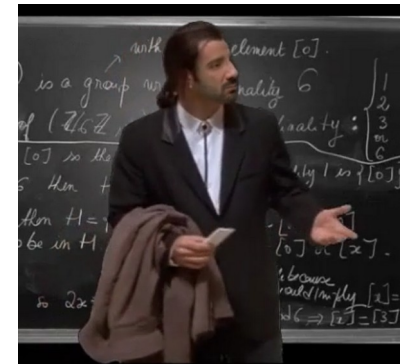
And... for accuracy?  
... Logistic Regression!

Outcome = 0,1

$$\text{Odds} = \frac{P(Y=1)}{1 - P(Y=1)}$$

$$\ln \left( \frac{P(Y=1)}{1 - P(Y=1)} \right) = \beta_0 + \beta_1 X_i$$

Log odds



And... for accuracy?  
... Logistic Regression!

Outcome = 0,1

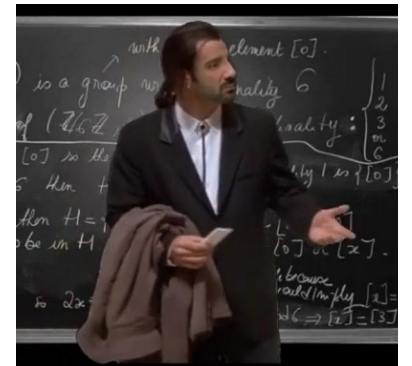
$$Odds = \frac{P(Y=1)}{1 - P(Y=1)}$$

$$\ln \left( \frac{P(Y=1)}{1 - P(Y=1)} \right) = \beta_0 + \beta_1 X_i$$

Log odds

$$P(Y) = \frac{1}{1 + e^{-(\beta_0 + \beta_j X_{ji})}}$$

Base of natural logarithm



## And... for accuracy? ... Logistic Regression!

Call:  
`glm(formula = rec_acc ~ PE, family = binomial, data = df)`

Deviance Residuals:

| Min     | 1Q      | Median | 3Q     | Max    |
|---------|---------|--------|--------|--------|
| -1.6697 | -1.1851 | 0.8432 | 0.9611 | 1.4105 |

Coefficients:

|             | Estimate | Std. Error | z value | Pr(> z )   |
|-------------|----------|------------|---------|------------|
| (Intercept) | 0.28799  | 0.02249    | 12.80   | <2e-16 *** |
| PE          | 0.84189  | 0.04397    | 19.15   | <2e-16 *** |

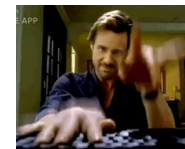
---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 11939 on 8999 degrees of freedom  
Residual deviance: 11819 on 8998 degrees of freedom  
AIC: 11823

Lmmworkshop.Rmd  
````{r log reg sim}`



The higher the
PE, the higher the
log odds of being
accurate.

And... for accuracy? ... Logistic Regression!

Call:
`glm(formula = rec_acc ~ PE, family = binomial, data = df)`

Deviance Residuals:

| Min | 1Q | Median | 3Q | Max |
|---------|---------|--------|--------|--------|
| -1.6697 | -1.1851 | 0.8432 | 0.9611 | 1.4105 |

Coefficients:

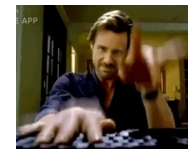
| | Estimate | Std. Error | z value | Pr(> z) |
|-------------|----------|------------|---------|------------|
| (Intercept) | 0.28799 | 0.02249 | 12.80 | <2e-16 *** |
| PE | 0.84189 | 0.04397 | 19.15 | <2e-16 *** |

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 11939 on 8999 degrees of freedom
Residual deviance: 11819 on 8998 degrees of freedom
AIC: 11823

Lmmworkshop.Rmd
````{r log reg sim}`



$$e^{\beta} = 2.32$$

For a unit increase in PE,  
participants' are 2.32 times more  
likely to be accurate. (odds of being  
correct are 2.32 times higher than  
those odds of being incorrect).

## And... for accuracy? ... Logistic Regression!

Call:  
`glm(formula = rec_acc ~ PE, family = binomial, data = df)`

Deviance Residuals:

| Min     | 1Q      | Median | 3Q     | Max    |
|---------|---------|--------|--------|--------|
| -1.6697 | -1.1851 | 0.8432 | 0.9611 | 1.4105 |

Coefficients:

|             | Estimate | Std. Error | z value | Pr(> z )   |
|-------------|----------|------------|---------|------------|
| (Intercept) | 0.28799  | 0.02249    | 12.80   | <2e-16 *** |
| PE          | 0.84189  | 0.04397    | 19.15   | <2e-16 *** |

---

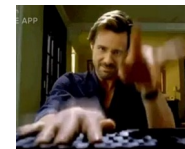
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 11939 on 8999 degrees of freedom  
Residual deviance: 11819 on 8998 degrees of freedom  
AIC: 11823

Lmmworkshop.Rmd

```
```{r log reg sim}```
```



Generalized Linear Mixed-effects model

Lmmworkshop.Rmd

```
``{r run generalised linear mixed-effects model}``
```



Generalized Linear Mixed-effects model – categorical predictor

```
Lmmworkshop.Rmd  
``{r run generalised linear mixed-effects model  
categorical}  
``
```



Difference ways of testing significance of fixed effects

`summary(model)` = It estimates the parameters and tests their significance through MLE or REML (the second is preferred)

`anova(model)` = It gives analysis of variance (F value). It does not provide significance test for GLMMs.

`car::Anova(model)` = It uses Analysis of deviance using Wald chi square test. It is good when you have several predictors and interactions. It works for GLMMs.

Specify the sum of squares:

Type II, when there is no significant interaction (main effects are evaluated after other main effects).

Type III in the presence of a significant interaction (main effects are tested after other main effects and interactions).

`anova(model1, model2)` = It computes LRT between two nested models. This procedure can lead to type I errors (Luke, 2017).

Troubleshooting: Convergence / Singularity

1. If you have continuous predictors, center or scale them.
2. Try a different optimizer. E.g, `control=glmerControl(optimizer="bobyqa")`
3. Increase the number of iterations.
`control=glmerControl(optimizer="bobyqa",optCtrl=list(maxfun=100000))` – default is 10000
4. Try lots of different optimizers with `allFit(yourMod)`. See which optimizer works and which not.
5. Use nlme.
6. Simplify models.
7. Use Bayesian methods with normalizing priors (MCMCglmm, brms)

For singularity: many authors do not worry about it.

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.
- Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of memory and language*, 59(4), 434-446.
- Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior research methods*, 49(4), 1494-1502.
- 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.
- Meteyard, L., & Davies, R. A. (2020). Best practice guidance for linear mixed-effects models in psychological science. *Journal of Memory and Language*, 112, 104092.

APA style Journal Article Reporting Standards:

(for quantitative studies, it can be found here)

<https://apastyle.apa.org/jars/quant-table-1.pdf>

- Provide information about exact p values (if NHST methods were used), effect-size estimates and confidence intervals, details of the models estimated, associated variance-covariance matrix or matrices, statistical software used (and packages), potential estimation problems (e.g., failure to converge).

Example of reporting:

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

"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 where random intercepts were included. Finally, a model with random intercept and slopes was tested. To test the significance of each random effect, we used likelihood ratio test (LRT) for each pairwise comparison, with $\alpha_{\text{LRT}} = 0.2$ (Matshucheck et al., 2017)"

For LMM, "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)".

For GLMM, "To test fixed effect, models were estimated with ML and Laplace Approximation" or " were estimated using parameter bootstrapping, to provide robust estimations (Luke, 2016)."

| | | | | | OR | | |
|-------------------------|----------|---------|--------|------|-------------|--------|-------|
| <i>Fixed effects</i> | Estimate | SE | z | p | e^{β} | 95% CI | |
| Intercept | -0.866 | 0.404 | -2.141 | .032 | 0.421 | 0.176 | 0.935 |
| PElevel2 | 1.084 | 0.369 | 2.94 | .003 | 2.958 | 1.427 | 6.554 |
| PElevel3 | 2.75 | 0.881 | 3.120 | .002 | 15.638 | 2.68 | 99.57 |
| | | | | | | | |
| <i>Random Effects</i> | Variance | St.dev. | Corr | | | | |
| Subjects
(Intercept) | 4.236 | 2.058 | | | | | |
| PElevel2 | 3.389 | 1.841 | -0.86 | | | | |
| PElevel3 | 22.146 | 4.706 | -0.81 | 0.99 | | | |