Day 2

- Multilevel Equations
- Steps of a Multilevel Analysis
- Maximum Approach
- Methodological Considerations
    - Centering
    - Small Level 2 N

Multilevel Equations

Multilevel Equations

- We are going to look at two ways of writing multilevel models.
  - Regression notation (used for articles and writing analysis code).
  - Distribution notation (more similar to how the output is presented; more intuitive)

- Introduce both notations using an example.
  - Students nested in classes
  - We are going to predict the student-level outcome *popularity* using
    - the student-level predictor *extraversion* and
    - class level predictor 'number of years of experience of the teacher' (*texp*).

Regression Notation

Regression Notation

- We start simple and just consider *popularity* and *extraversion*.

- Regular linear regression:

$$popularity_{ij} = b_{intercept} + b_{ext} extraversion_{ij} + e_{ij}$$

- Multilevel regression:

$$popularity_{ij} = b_{intercept\_j} + b_{ext\_j} extraversion_{ij} + e_{ij}$$

Regression Notation

- We start simple and just consider *popularity* and *extraversion*.

- Regular linear regression:

$$popularity_{ij} = b_{intercept} + b_{ext} extraversion_{ij} + e_{ij}$$

- Multilevel regression:

$$popularity_{ij} = b_{intercept\_j} + b_{ext\_j} extraversion_{ij} + e_{ij}$$

Regression Notation

$$popularity_{ij} = b_{intercept\_j} + b_{ext\_j} extraversion_{ij} + e_{ij}$$

- So the coefficients can vary across classes (this is why multilevel models are also called random coefficient models).

- Since we have variation on two levels, student (level 1) and class (level 2), let's explicitly incorporate that into the equation.

Regression Notation

Level 1: $popularity_{ij} = \ b_{intercept\_j} + b_{ext\_j} extraversion_{ij} + e_{ij}$

Level 2:

Regression Notation

Level 1: $popularity_{ij} = b_{intercept\_j} + b_{ext\_j}extraversion_{ij} + e_{ij}$

Regular regression

Level 2: $b_{intercept\_j} = \gamma_{intercept\_intercept}$

$b_{ext\_j} = \gamma_{intercept\_ext}$

Regression Notation

Level 1: $popularity_{ij} = \ b_{intercept\_j} + b_{ext\_j}extraversion_{ij} + e_{ij}$

Multilevel regression with
random intercept
(with no level 2 predictors)

Level 2: $b_{intercept\_j} = \gamma_{intercept\_intercept} + u_{intercept\_j}$

$b_{ext\_j} = \gamma_{intercept\_ext}$

Regression Notation

Level 1: $popularity_{ij} = \ b_{intercept\_j} + b_{ext\_j} extraversion_{ij} + e_{ij}$

Multilevel regression with random intercept and slope (with no level 2 predictors)

Level 2: $b_{intercept\_j} = \gamma_{intercept\_intercept} + u_{intercept\_j}$

$b_{ext\_j} = \gamma_{intercept\_ext} + u_{ext\_j}$

Regression Notation

- Let's add in the main effect of *teacher experience* (level 2)

Level 1: $popularity_{ij} = b_{intercept\_j} + b_{ext\_j} extraversion_{ij} + e_{ij}$

Level 2: $b_{intercept\_j} = \gamma_{intercept\_intercept} + \gamma_{texp\_intercept} * texp_j + u_{intercept\_j}$

$b_{ext\_j} = \gamma_{intercept\_ext} + u_{ext\_j}$

Regression Notation

- Let's add in the main effect of *teacher experience* (level 2)

Level 1:

$$popularity_{ij} = b_{intercept\_j} + b_{ext\_j} extraversion_{ij} + e_{ij}$$

Level 2:

$$b_{intercept\_j} = \gamma_{intercept\_intercept} + \gamma_{texp\_intercept} * texp_j + u_{intercept\_j}$$

$$b_{ext\_j} = \gamma_{intercept\_ext} + u_{ext\_j}$$

Regression Notation

- Of course, *texp* might also explain between-class differences in the effect of *extraversion* on *popularity*

Level 1:

$$popularity_{ij} = \ b_{intercept\_j} + b_{ext\_j} extraversion_{ij} + e_{ij}$$

Multilevel regression with random intercept and slope and cross-level interaction

Level 2:

$$b_{intercept\_j} = \gamma_{intercept\_intercept} + \gamma_{texp\_intercept} * texp_j + u_{intercept\_j}$$

$$b_{ext\_j} = \gamma_{intercept\_ext} + \gamma_{texp\_ext} * texp_j + u_{ext\_j}$$

Tilburg University

tesc

Regression Notation

Level 1 :

$$popularity_{ij} = b_{intercept\_j} + b_{ext\_j} extraversion_{ij} + e_{ij}$$

Level 2:

$$b_{intercept\_j} = \gamma_{intercept\_intercept} + \gamma_{texp\_intercept} * texp_j + u_{intercept\_j}$$

$$b_{ext\_j} = \gamma_{intercept\_ext} + \gamma_{texp\_ext} * texp_j + u_{ext\_j}$$

Level 1 + 2:

$$popularity_{ij} = \gamma_{intercept\_intercept} + \gamma_{texp\_intercept} * texp_j + \gamma_{intercept_{ext}} * extraversion_{ij} + \gamma_{texp\_ext} * texp_j * extraversion_{ij} + e_{ij} + u_{intercept\_j} + u_{ext\_j} * extraversion_{ij}$$

Tilburg University

tesc

Regression Notation

Level 1 :

$$popularity_{ij} = b_{intercept\_j} + b_{ext\_j} extraversion_{ij} + e_{ij}$$

Level 2:

$$b_{intercept\_j} = \gamma_{intercept\_intercept} + \gamma_{texp\_intercept} * texp_j + u_{intercept\_j}$$

$$b_{ext\_j} = \gamma_{intercept\_ext} + \gamma_{texp\_ext} * texp_j + u_{ext\_j}$$

Level 1 + 2:

$$popularity_{ij} = \gamma_{intercept\_intercept} + \gamma_{texp\_intercept} * texp_j + \gamma_{intercept_{ext}} * extraversion_{ij} + \gamma_{texp\_ext} * texp_j * extraversion_{ij} + e_{ij} + u_{intercept\_j} + u_{ext\_j} * extraversion_{ij}$$

Tilburg University

tesc

Regression Notation

Level 1 :

$$popularity_{ij} = b_{intercept\_j} + b_{ext\_j} extraversion_{ij} + e_{ij}$$

Level 2:

$$b_{intercept\_j} = \gamma_{intercept\_intercept} + \gamma_{texp\_intercept} * texp_j + u_{intercept\_j}$$

$$b_{ext\_j} = \gamma_{intercept\_ext} + \gamma_{texp\_ext} * texp_j + u_{ext\_j}$$

Level 1 + 2:

$$popularity_{ij} = \gamma_{intercept\_intercept} + \gamma_{texp\_intercept} * texp_j + \gamma_{intercept_{ext}} * extraversion_{ij} + \gamma_{texp\_ext} * texp_j * extraversion_{ij} + e_{ij} + u_{intercept\_j} + u_{ext\_j} * extraversion_{ij}$$

Regression Notation

Level 1 + 2:

$$popularity_{ij} =$$

$$\overbrace{\gamma_{intercept\_intercept}}^{overall\ intercept} + \overbrace{\gamma_{intercept\_ext}}^{overall\ effect\ of\ extraversion} * extraversion_{ij} + \overbrace{\gamma_{texp\_intercept}}^{main\ effect\ of\ texp} * texp_j + \overbrace{\gamma_{texp\_ext}}^{cross-level\ interaction} * texp_j * extraversion_{ij}$$

$$+ \overbrace{e_{ij}}^{level\ 1\ residual} + \overbrace{u_{intercept\_j} + u_{ext\_j} * extraversion_{ij}}^{level\ 2\ residuals}$$

TILBURG UNIVERSITY

tesc

Regression Notation

Level 1 :

$$popularity_{ij} = b_{intercept\_j} + b_{ext\_j} extraversion_{ij} + \overbrace{e_{ij}}^{level\ 1\ residuals}$$

Level 2:

$$b_{intercept\_j} = \overbrace{\gamma_{intercept\_intercept}}^{overall\ intercept} + \overbrace{\gamma_{texp\_intercept}}^{main\ effect\ of\ texp} * texp_j + \overbrace{u_{intercept_j}}^{level\ 2\ residuals}$$

$$b_{ext\_j} = \overbrace{\gamma_{intercept\_ext}}^{overall\ effect\ of\ ext} + \overbrace{\gamma_{texp\_ext}}^{cross-level\ interaction} * texp_j + \overbrace{u_{ext\_j}}^{level\ 2\ residual}$$

TILBURG UNIVERSITY

tesc

Regression Notation

Level 1 : $Y_{ij} = b_{intercept\_j} + \sum_{p=1}^{P} b_{p\_j} X_{p,ij} + e_{ij}$

Level 2: $b_{intercept\_j} = \gamma_{intercept\_intercept} + \sum_{q=1}^{Q} \gamma_{q\_intercept} * Z_{q,j} + u_{intercept\_j}$

$b_{p\_j} = \gamma_{intercept\_p} + \sum_{q=1}^{Q} \gamma_{q\_p} * Z_{q,j} + u_{p\_j}$

Distribution Notation

Distribution Notation

Level 1: $popularity_{ij} \sim Normal(mean, sd)$

Level 2: $b_{intercept\_j}$

b_{ext_j}

Tilburg University

tesc

Distribution Notation

Level 1: $popularity_{ij} \sim Normal(b_{intercept\_j} + b_{ext\_j} * extraversion_{ij},\ \sigma_e)$

Level 2: $b_{intercept\_j}$

b_{ext_j}

Distribution Notation

Level 1: $popularity_{ij} \sim Normal\left(b_{intercept\_j} + b_{ext\_j} * extraversion_{ij},\ \sigma_e\right)$

Level 2: $b_{intercept\_j} \sim Normal(mean, sd)$

$b_{ext\_j} \sim Normal(mean, sd)$

Distribution Notation

Level 1: $popularity_{ij} \sim Normal(b_{intercept\_j} + b_{ext\_j} * extraversion_{ij},\ \sigma_e)$

Level 2: $b_{intercept\_j} \sim Normal(\gamma_{intercept\_intercept}, 0)$

$b_{ext\_j} \sim Normal(\gamma_{intercept\_ext}, 0)$

Regular regression

Distribution Notation

Level 1: $popularity_{ij} \sim Normal(b_{intercept\_j} + b_{ext\_j} * extraversion_{ij}, \sigma_e)$

Level 2: $b_{intercept\_j} \sim Normal\left(\gamma_{intercept\_intercept}, \sigma_{intercept}\right)$

$b_{ext\_j} \sim Normal\left(\gamma_{intercept\_ext}, \sigma_{ext}\right)$

Multilevel regression with random intercept and slope (with no level 2 predictors)

Distribution Notation

Level 1:

$$popularity_{ij} \sim Normal(b_{intercept\_j} + b_{ext\_j} * extraversion_{ij},\ \sigma_e)$$

Level 2:

$$b_{intercept\_j} \sim Normal\left(\gamma_{intercept\_intercept} + \gamma_{texp\_intercept} * texp_j\,, \sigma_{intercept}\right)$$

$$b_{ext\_j} \sim Normal\left(\gamma_{intercept\_ext}\,, \sigma_{ext}\right)$$

Multilevel regression with random intercept and slope and main effect of level 2 predictors

TILBURG UNIVERSITY

tesc

Distribution Notation

Level 1:

$$popularity_{ij} \sim Normal(b_{intercept\_j} + b_{ext\_j} * extraversion_{ij},\ \sigma_e)$$

Level 2:

$$b_{intercept\_j} \sim Normal\left(\gamma_{intercept\_intercept} + \gamma_{texp\_intercept} * texp_j, \sigma_{intercept}\right)$$

$$b_{ext\_j} \sim Normal\left(\gamma_{intercept\_ext} + \gamma_{texp\_ext} * texp_j, \sigma_{ext}\right)$$

Multilevel regression with random intercept and slope and main effect of level 2 predictors and cross-level interaction

Distribution Notation

Level 1: $Y_{ij} \sim Normal\left(b_{intercept\_j} + \sum_{p=1}^{P} b_{p\_j}\, X_{p,ij},\ \sigma_e\right)$

Level 2: $b_{intercept\_j} \sim Normal\left(\gamma_{intercept\_intercept} + \sum_{q=1}^{Q} \gamma_{q\_intercept} * Z_{q,j},\ \sigma_{intercept}\right)$

$b_{p\_j} \sim Normal\left(\gamma_{intercept\_p} + \sum_{q=1}^{Q} \gamma_{q\_p} * Z_{q,j},\ \sigma_p\right)$

Steps of a Multilevel Analysis

Steps of a multilevel analysis

1. Check whether multilevel is necessary
2. Add first-level variables
3. Add second-level variables
4. Add random slopes
5. Add cross-level interactions (if random slopes are present)

Step 1: Check whether multilevel is necessary

Step 1: Check whether multilevel is necessary

Analyze a two-level model with no explanatory variables: the intercept-only model

- This means there are no predictors—just an intercept that varies across groups.
- The model has a fixed intercept with between-group variance (i.e., the means of the outcome variable are allowed to differ across groups)

- If multilevel modeling isn't needed (i.e., there is no difference in the errors across groups), we can combine data from all groups and treat it as one large sample, using regular regression!

Step 1: Check whether multilevel is necessary

- Hypotheses
  - H_0: individuals in the same group aren't more alike than individuals in different groups; the level-2 variance (in the intercepts) is zero
  - H_1: individuals in the same group are more alike than individuals in different groups; the level-2 variance (in the intercepts) is larger than zero

Step 1: Check whether multilevel is necessary

- We have a measure of whether multilevel analysis is necessary, the intra-class correlation, ICC
- ICC:
  - The percentage variance at the second level; or:
  - The expected correlation between the observations on the dependent variable of two randomly chosen units that are in the same group.
  - $\rho = \frac{\sigma^2_{intercept}}{\sigma^2_{intercept}+\sigma^2_e}$, where

    $\sigma^2_{intercept}$ is variance of the second level errors ("unexplained variance at the second level"), and

    σ^2_e is variance of the first level errors ("unexplained variance at the first level")

TILBURG UNIVERSITY

tesc

Step 1 in R

- The data is a subsample from the 1982 High School and Beyond Survey
- The data file, called hsb, consists of 7185 students nested in 160 schools
- The outcome variable of interest is the student-level (level 1) math achievement score (mathach)
- The variable ses is the socio-economic status of a student and therefore is at the student level
- The variable sector is an indicator variable indicating if a school is public or catholic and is therefore a school-level variable
    - There are 90 public schools (sector=0) and 70 catholic schools (sector=1) in the sample.

Step 1 in R

```
> # Check whether multilevel is necessary (using intercept-only models)
> IO <- lm(mathach ~ 1, hsb)
> IO_ML <- lmer(mathach ~ 1 + (1 | id), hsb)
> anova(IO_ML, IO)
refitting model(s) with ML (instead of REML)
Data: hsb
Models:
IO: mathach ~ 1
IO_ML: mathach ~ 1 + (1 | id)
      npar   AIC   BIC logLik deviance  Chisq Df Pr(>Chisq)
IO       2 48104 48117 -24050    48100
IO_ML    3 47122 47142 -23558    47116 983.92  1  < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Allowing for between-group differences in the intercepts significantly improves model fit

Step 1 in R

```
> summary(IO_ML)
Linear mixed model fit by REML ['lmerMod']
Formula: mathach ~ 1 + (1 | id)
   Data: hsb

REML criterion at convergence: 47116.8

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.0631 -0.7539  0.0267  0.7606  2.7426

Random effects:
 Groups   Name        Variance Std.Dev.
 id       (Intercept)  8.614   2.935
 Residual             39.148   6.257
Number of obs: 7185, groups:  id, 160

Fixed effects:
            Estimate Std. Error t value
(Intercept)  12.6370     0.2444   51.71
>
> # "Variance id" / ("Variance residual" + "Variance id")
> 8.614/(8.614+39.148) # ICC = 0.1803526
[1] 0.1803526
> # --> variance explained by school membership
```

Together they make up the total variance: 8.614 of the variance is at the school level and 39.148 is the variance within schools

Average achievement score across all schools and students

Step 2: Add first-level variables

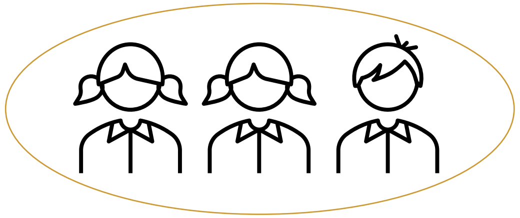
Step 2: Add first-level variables

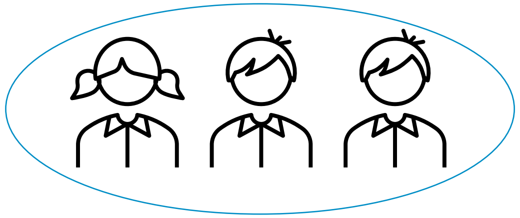
- We want to check if individual-level variables (level 1) explain within-group variation in the outcome
  - Add all first-level variables (+ their interactions) and check whether they are significant or not

- Hypotheses:
  - H_0: There is no relation between the dependent and the explanatory variable(s)
  - H_1: There is a (positive/negative) relation between the dependent and the explanatory variable(s)

Step 2: Add first-level variables

- Calculate the explained variance at both level 1 and at level 2.
  - Variables at the lowest level can explain variance at the first level (e.g., within schools)
    - Girls are more popular than boys. 😎
  - Variables at the lowest levels can explain variance at the higher levels
    - Girls are more popular than boys. 😎
    - In a class with a lot of girls, the mean popularity is higher: the distribution of a level-1 predictor may differ across classes and thus the mean popularity would be different. However, this difference is explained by the difference in amounts of girls.
  - We have multiple levels on which variance can be explained!





Step 2 in R

```
> # Step 2: Add first-level variables
> ML_level1 <- lmer(mathach ~ 1 + ses  + (1 | id), hsb)
> summary(ML_level1)
Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']
Formula: mathach ~ 1 + ses + (1 | id)
   Data: hsb

REML criterion at convergence: 46645.2

Scaled residuals:
     Min       1Q   Median       3Q      Max
-3.12607 -0.72720  0.02188  0.75772  2.91912

Random effects:
 Groups   Name        Variance Std.Dev.
 id       (Intercept)  4.768   2.184
 Residual             37.034   6.086
Number of obs: 7185, groups:  id, 160

Fixed effects:
             Estimate Std. Error        df t value Pr(>|t|)
(Intercept)   12.6575     0.1880  148.3022   67.33   <2e-16 ***
ses            2.3902     0.1057 6838.0776   22.61   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
    (Intr)
ses 0.003
```

There is a significant effect of *ses* on achievement

Step 2 in R

```
> # Now, we calculate the explained variances on Level 1 and 2 by subtracting the
> # residual variance of the large model from the residual variance of the smaller
> # model (see "vcov" in the output) and dividing it by the variance of the
> # smaller (i.e., initial) model
> VarianceIO <- as.data.frame(VarCorr(IO_ML))
> VarianceLv1 <- as.data.frame(VarCorr(ML_level1))
>
> VarianceIO
       grp        var1 var2      vcov    sdcor
1       id (Intercept) <NA>  8.614025 2.934966
2 Residual        <NA> <NA> 39.148322 6.256862
> VarianceLv1
       grp        var1 var2      vcov    sdcor
1       id (Intercept) <NA>  4.768175 2.183615
2 Residual        <NA> <NA> 37.034399 6.085589
>
> # Explained Variance on Level 1
> (VarianceIO[2, 4] - VarianceLv1[2, 4]) / VarianceIO[2, 4]
[1] 0.0539978
> # 0.0539978
>
> # Explained Variance on Level 2
> (VarianceIO[1, 4] - VarianceLv1[1, 4]) / VarianceIO[1, 4]
[1] 0.4464638
> # 0.4464638
```

Together they make up the total (unexplained) variance

Together they make up what unexplained variance is left after adding *ses* as a predictor

Although *ses* is a level-1 variable, it mainly explains variance on the second level, the school level: in a school with a high *ses* average, achievement is better.

Step 3: Add second-level variables

Step 3: Second-level variables

- We want to check if group-level variables (level 2) can explain between-group variation in the outcome
  - Add all second-level variables (+ their interactions) and check whether they are significant or not

- Hypotheses:
  - H_0: There is no relation between the explanatory variable(s) and the mean score of the dependent variable
  - H_1: There is a (positive/negative) relation between the explanatory variable(s) and the mean score of the dependent variable

TILBURG UNIVERSITY

tesc

Step 3 in R

```
> # Step 3: Add second-level variables
>
> ML_level1_2 <- lmer(mathach ~ 1 + ses  + (1 | id) + catholic, hsb)
> summary(ML_level1_2)# the sector (catholic) is significantly related to mathach
Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']
Formula: mathach ~ 1 + ses + (1 | id) + catholic
   Data: hsb

REML criterion at convergence: 46611.2

Scaled residuals:
     Min       1Q   Median       3Q      Max
-3.14857 -0.73100  0.01929  0.75366  2.92634

Random effects:
 Groups   Name        Variance Std.Dev.
 id       (Intercept)  3.685   1.920
 Residual             37.037   6.086
Number of obs: 7185, groups:  id, 160

Fixed effects:
                 Estimate Std. Error        df t value Pr(>|t|)
(Intercept)       11.7189     0.2281  153.5842  51.386  < 2e-16 ***
ses                2.3747     0.1055 6738.8583  22.511  < 2e-16 ***
catholiccatholic   2.1008     0.3411  147.3574   6.159 6.64e-09 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
            (Intr) ses
ses          0.063
cathlccthlc -0.672 -0.091
```

There is a significant effect of the sector on achievement

Step 3 in R

```
> VarianceLV12 <- as.data.frame(VarCorr(ML_level1_2))
>
> VarianceLv1
       grp        var1 var2      vcov    sdcor
1       id (Intercept) <NA>  4.768175 2.183615
2 Residual        <NA> <NA> 37.034399 6.085589
> VarianceLV12
       grp        var1 var2     vcov    sdcor
1       id (Intercept) <NA>  3.68504 1.919646
2 Residual        <NA> <NA> 37.03691 6.085796
>
> # Explained Variance on Level 1 (compared to model with just level-1 predictors)
> (VarianceLv1[2, 4] - VarianceLV12[2, 4]) / VarianceLv1[2, 4]
[1] -6.788361e-05
> # 0. No explained variance since you only added a level 2 predictor
>
> # Explained Variance on Level 2 (compared to model with just level 1 predictors)
> (VarianceLv1[1, 4] - VarianceLV12[1, 4]) / VarianceLv1[1, 4]
[1] 0.2271592
> # 0.2271592
```

Compare to the step-2 model

No explained variance on level 1: a higher-level predictor can never explain variance on the lower level(s)!

Step 4: Add random slopes

Step 4: Add random slopes

- We want to check if the relationship between level 1 predictors and the outcome varies across groups (level 2).
- We allow the regression coefficients for the level 1 predictors to differ across the higher units.
  - We check whether the variance in these regression coefficients is significantly different from zero.

- Hypotheses:
  - H_0: The relation between the explanatory variable and the dependent variable is the same within all level-2 units (the variance is = 0)
  - H_1: The relation between the explanatory variable and the dependent variable is not the same within all level-2 units (the variance is > 0)

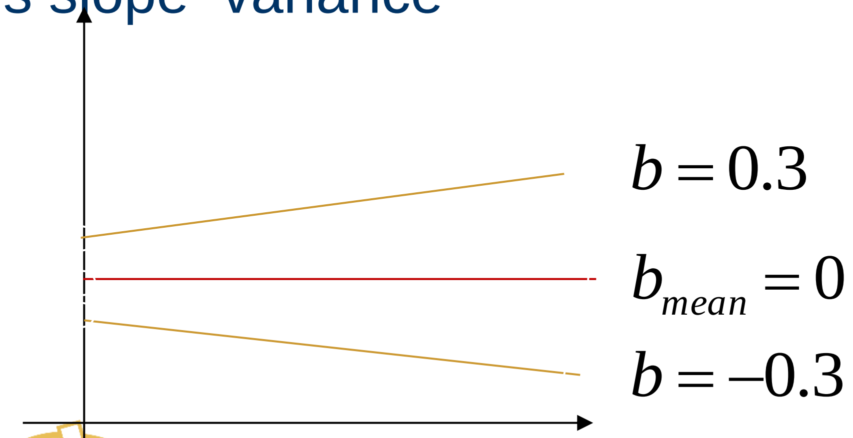
TILBURG UNIVERSITY

tesc

Step 4: Add random slopes

- We test for random slope variance variable by variable:
  - Add all significant variables from step 2 and make them random one by one (keep the ones random that have significant variance across groups)

- Once this is done, also consider adding variables that were omitted in step 2: add them with a random effect one by one:
  - It is possible that an explanatory variable has no significant mean regression slope, but that there is slope variance



Step 4 in R

```
> ML_level1_2_RE <- lmer(mathach ~ 1 + ses  + (1 + ses | id) + catholic, hsb)
> anova(ML_level1_2_RE, ML_level1_2)
refitting model(s) with ML (instead of REML)
Data: hsb
Models:
ML_level1_2: mathach ~ 1 + ses + (1 | id) + catholic
ML_level1_2_RE: mathach ~ 1 + ses + (1 + ses | id) + catholic
               npar   AIC   BIC logLik deviance  Chisq Df Pr(>Chisq)
ML_level1_2       5 46616 46651 -23303    46606
ML_level1_2_RE    7 46611 46660 -23299    46597 9.0438  2    0.01087 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>
> # you would do this per lower level variable;
> # --> here we have only one and adding a random slope for it significantly
> #     improves model fit.
```

Step 4 in R

```
> summary(ML_level1_2_RE)
Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']
Formula: mathach ~ 1 + ses + (1 + ses | id) + catholic
   Data: hsb

REML criterion at convergence: 46601.9

Scaled residuals:
    Min      1Q  Median      3Q     Max 
-3.1373 -0.7296  0.0225  0.7568  2.8920 

Random effects:
 Groups   Name        Variance Std.Dev. Corr
 id       (Intercept)  3.9646  1.991        
          ses          0.4343  0.659    0.55
 Residual             36.8008  6.066        
Number of obs: 7185, groups:  id, 160

Fixed effects:
                 Estimate Std. Error       df t value Pr(>|t|)    
(Intercept)       11.4729     0.2315 153.7963  49.568  < 2e-16 ***
ses                2.3854     0.1179 157.8423  20.238  < 2e-16 ***
catholiccatholic   2.5408     0.3445 151.3087   7.375 1.01e-11 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
            (Intr) ses   
ses          0.228       
cathlccthlc -0.655 -0.079
```

Step 5: Add cross-level interactions

Step 5: Add cross-level interactions

- We want to see if the variance in slopes across level-2 groups can be explained by level-2 predictors.
- Add cross-level interactions between level-1 explanatory variables that had significant slope variation and level-2 explanatory variables

- Hypotheses:
  - H_0: The explanatory variable cannot explain the variance in the relations between the explanatory variable and the dependent variable in different level-2 units
  - H_1: the explanatory variable explains (a part of) the variance in the relations between the explanatory variable and the dependent variable in different level-2 units

Step 5 in R

```
> # Step 5: Add cross-level interactions
> ML_level1_2_RE_CL <- lmer(mathach ~ 1 + ses * catholic  + (1 + ses | id), hsb)
Warning message:
In checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv,  :
  Model failed to converge with max|grad| = 0.00578927 (tol = 0.002, component 1)
> #--> too complex ("Model failed to converge"); let's use brm instead
> ML_level1_2_RE_CL <- brm(mathach ~ 1 + ses * catholic  + (1 + ses | id), hsb)
Compiling Stan program...
Start sampling
```

Step 5 in R

```
> summary(ML_level1_2_RE_CL)
 Family: gaussian
  Links: mu = identity; sigma = identity
Formula: mathach ~ 1 + ses * catholic + (1 + ses | id)
   Data: hsb (Number of observations: 7185)
  Draws: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
         total post-warmup draws = 4000

Group-Level Effects:
~id (Number of levels: 160)
                   Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sd(Intercept)          1.97      0.15     1.70     2.28 1.00     1079     1461
sd(ses)                0.35      0.17     0.03     0.70 1.00      970     1459
cor(Intercept,ses)     0.57      0.32    -0.26     0.98 1.00     3099     2625

Population-Level Effects:
                     Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept               11.75      0.23    11.30    12.21 1.00     1410     2227
ses                      2.96      0.15     2.66     3.24 1.00     3813     2870
catholiccatholic         2.14      0.36     1.46     2.85 1.00     1319     1957
ses:catholiccatholic    -1.31      0.22    -1.75    -0.87 1.00     3860     3048

Family Specific Parameters:
      Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sigma     6.06      0.05     5.97     6.16 1.00     7255     2997

Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS
and Tail_ESS are effective sample size measures, and Rhat is the potential
scale reduction factor on split chains (at convergence, Rhat = 1).
>
> #--> Look at the credibility intervals: effects are significant if zero is not
> # included. Thus, the effect of ses on mathach is lower for catholic schools
> # than for public schools.
```

There is a significant interaction effect: the effect is negative and thus the effect of *ses* on *achievement* is lower for catholic schools than for public schools.

Maximum Approach

Maximum Approach

Think about the following questions:

- What are the possible advantages of starting with a full model?
- What are the potential downsides?

Maximum Approach

- Advantages of Starting with a Full Model:
  - We know that effects likely differ between people/units: immediately adding random slopes makes sense
  - If we initially omit random effect, we base intermediate decisions on misspecified models

- Disadvantages of Starting with a Full Model:
  - More complex models require more computational resources: if the model is too complex, you need to remove random effects (one by one) and then deciding which one may not be straightforward
  - More complex models can be harder to interpret

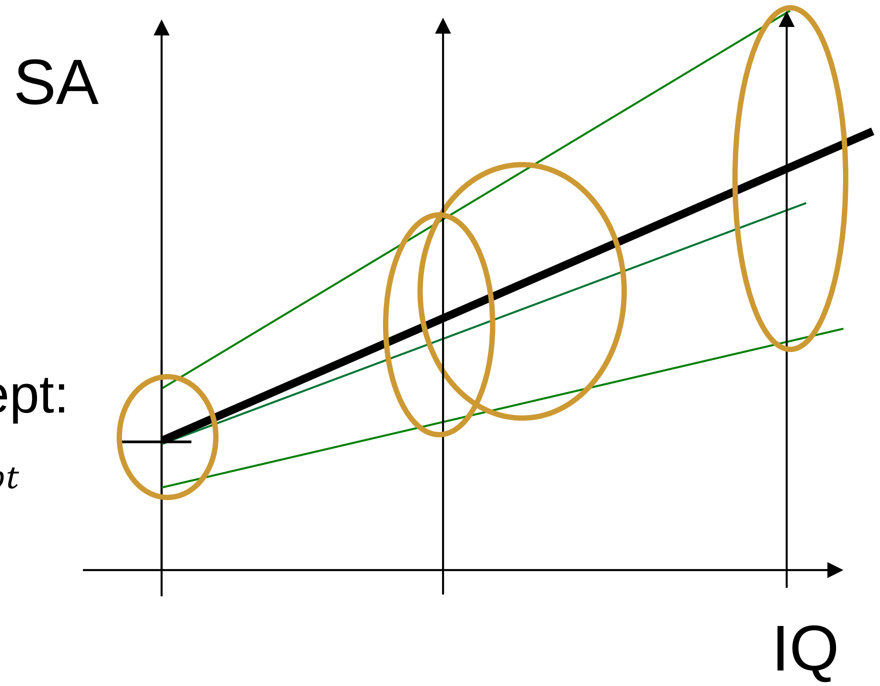
Methodological Considerations: Centering

TILBURG UNIVERSITY

tesc

Centering

Average Slope = $\gamma_{intercept\_IQ}$



Average Intercept:
$\gamma_{intercept\_intercept}$

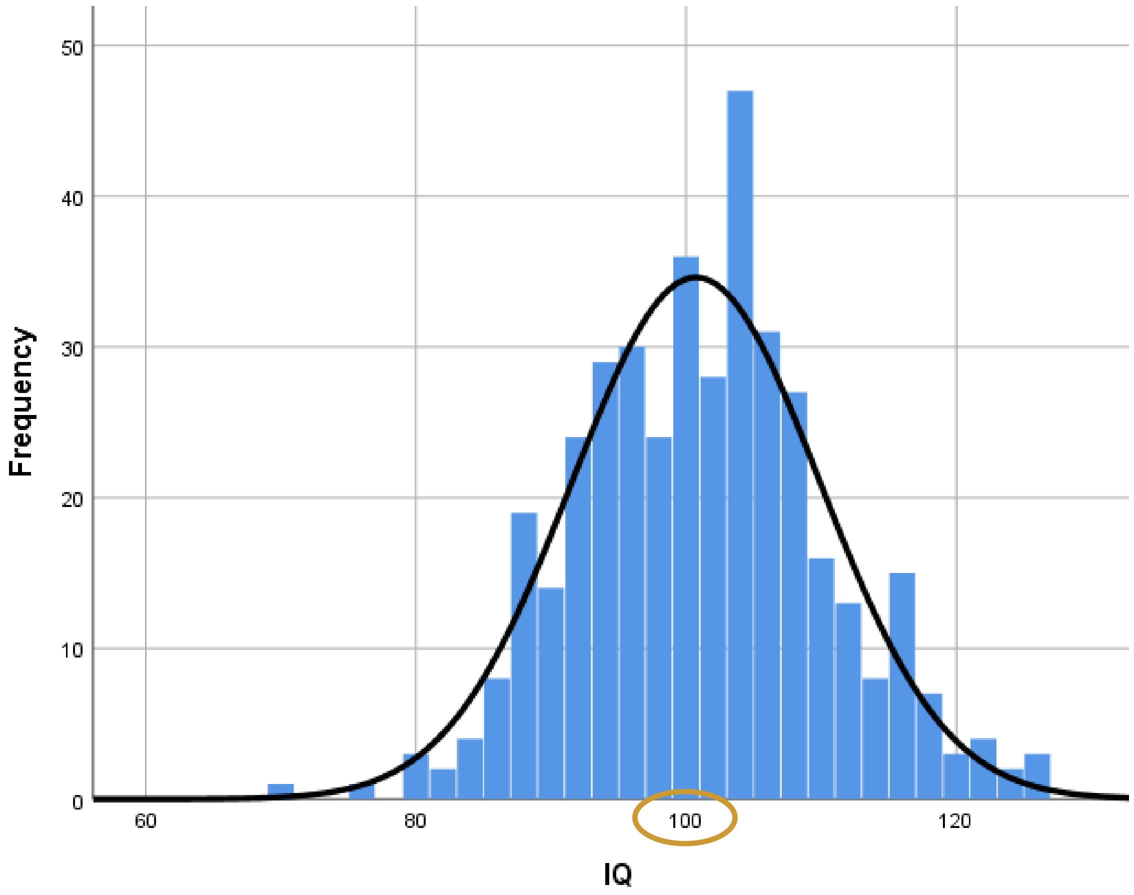
Class 1: Slope = $\gamma_{intercept\_IQ} + u_{IQ\_1}$

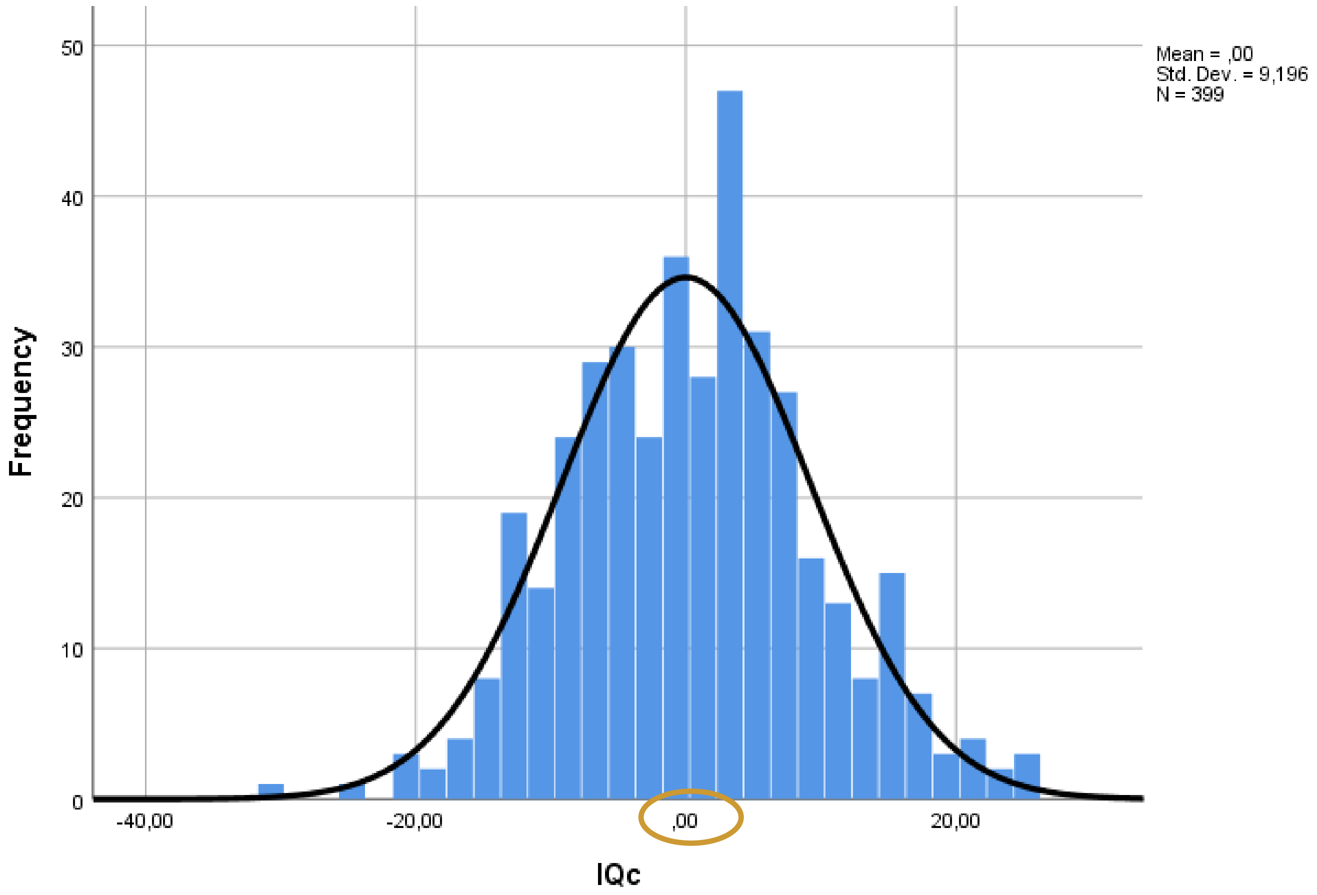
Class 2: Slope = $\gamma_{intercept\_IQ} + u_{IQ\_2}$

Class 3: Slope = $\gamma_{intercept\_IQ} + u_{IQ\_3}$

$$b_{\mathrm{intercept}\_j} = \gamma_{00} + \gamma_{01} Classsize_j + u_{intercept\_j}$$
$$b_{IQ\_j} = \gamma_{10} + u_{IQ\_j}$$

Centering

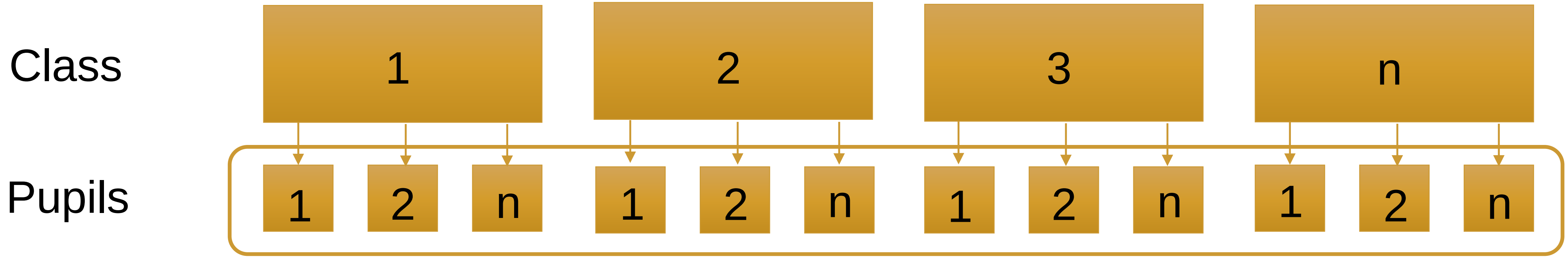


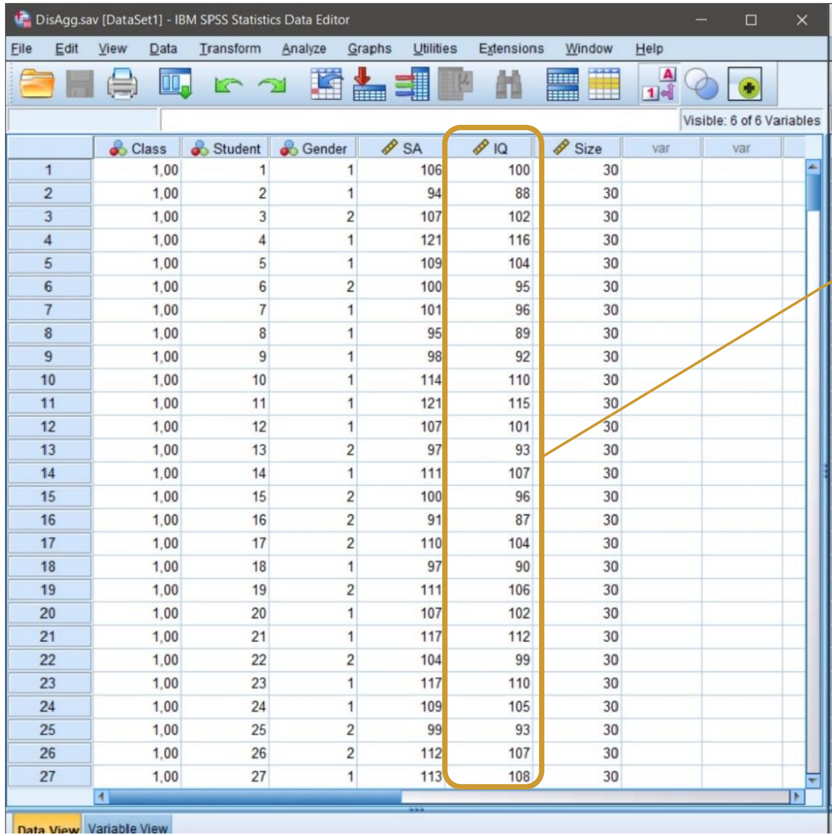


TILBURG ◆ UNIVERSITY

tesc

Centering





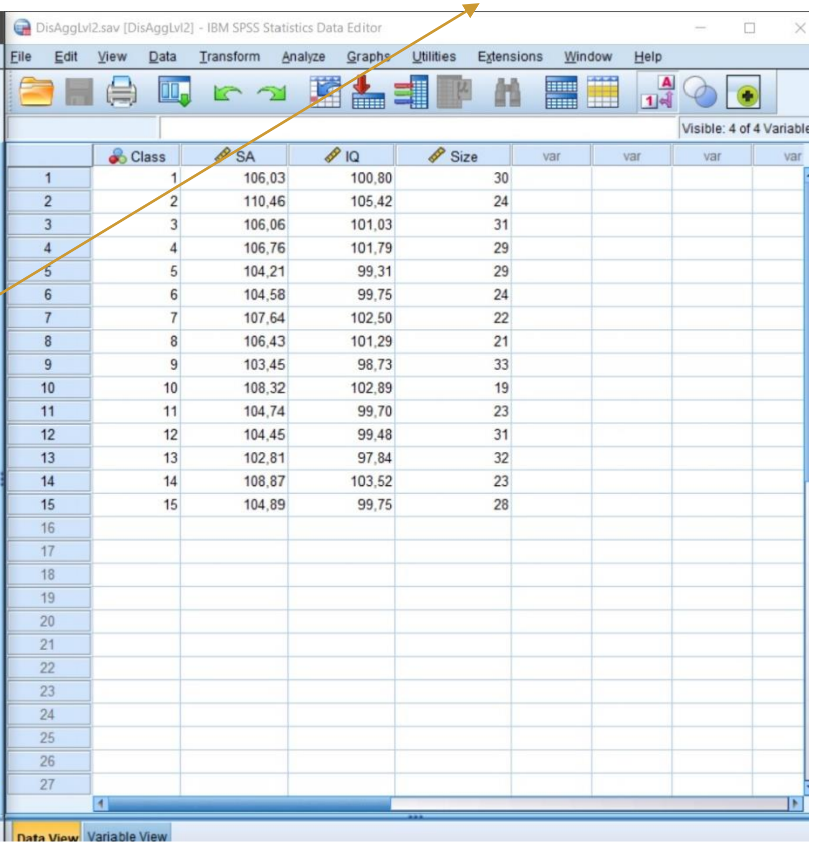
DisAgg.sav [DataSet1] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help

Visible: 6 of 6 Variables

	Class	Student	Gender	SA	IQ	Size
1	1,00	1	1	106	100	30
2	1,00	2	1	94	88	30
3	1,00	3	2	107	102	30
4	1,00	4	1	121	116	30
5	1,00	5	1	109	104	30
6	1,00	6	2	100	95	30
7	1,00	7	1	101	96	30
8	1,00	8	1	95	89	30
9	1,00	9	1	98	92	30
10	1,00	10	1	114	110	30
11	1,00	11	1	121	115	30
12	1,00	12	1	107	101	30
13	1,00	13	2	97	93	30
14	1,00	14	1	111	107	30
15	1,00	15	2	100	96	30
16	1,00	16	2	91	87	30
17	1,00	17	2	110	104	30
18	1,00	18	1	97	90	30
19	1,00	19	2	111	106	30
20	1,00	20	1	107	102	30
21	1,00	21	1	117	112	30
22	1,00	22	2	104	99	30
23	1,00	23	1	117	110	30
24	1,00	24	1	109	105	30
25	1,00	25	2	99	93	30
26	1,00	26	2	112	107	30
27	1,00	27	1	113	108	30

Data View Variable View



DisAggLvl2.sav [DisAggLvl2] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help

Visible: 4 of 4 Variable

	Class	SA	IQ	Size
1	1	106,03	100,80	30
2	2	110,46	105,42	24
3	3	106,06	101,03	31
4	4	106,76	101,79	29
5	5	104,21	99,31	29
6	6	104,58	99,75	24
7	7	107,64	102,50	22
8	8	106,43	101,29	21
9	9	103,45	98,73	33
10	10	108,32	102,89	19
11	11	104,74	99,70	23
12	12	104,45	99,48	31
13	13	102,81	97,84	32
14	14	108,87	103,52	23
15	15	104,89	99,75	28
16				
17				
18				
19				
20				
21				
22				
23				
24				
25				
26				
27				

Data View Variable View

Descriptive Statistics

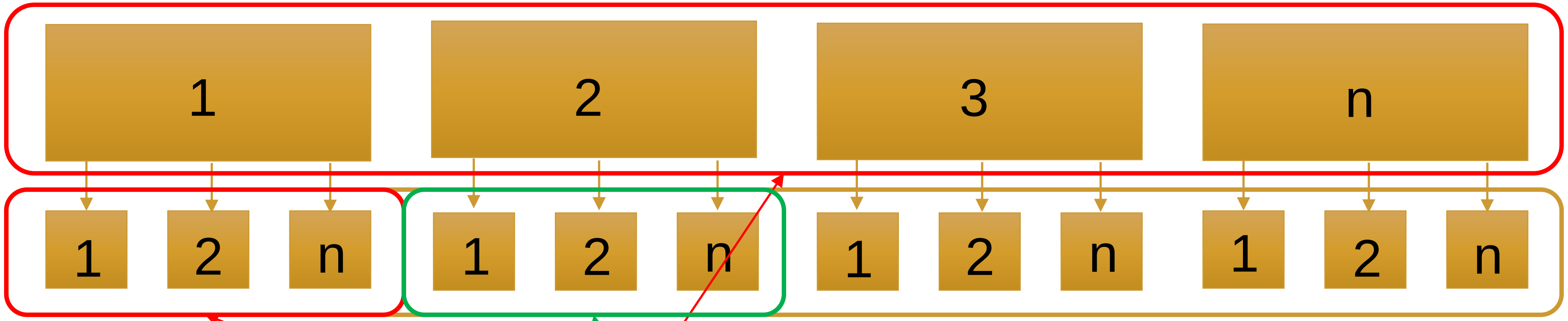
	N	Minimum	Maximum	Mean	Std. Deviation
IQ	399	70	126	100,72	9,196
Valid N (listwise)	399				

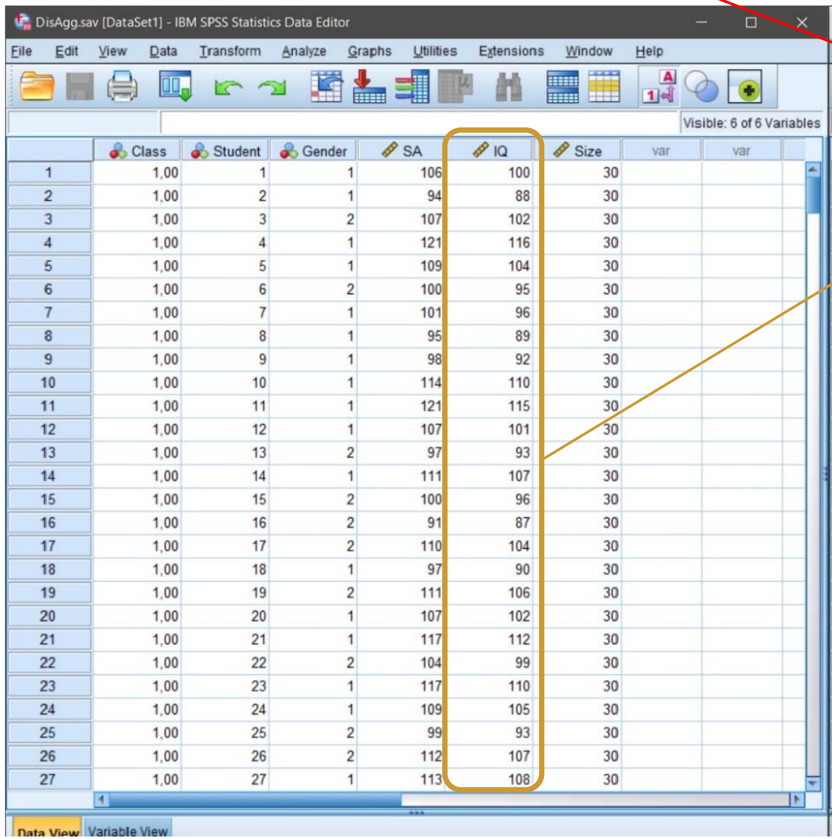
$$IQ_c = \boxed{IQ_{ij}} - \boxed{100.72}$$

Centering

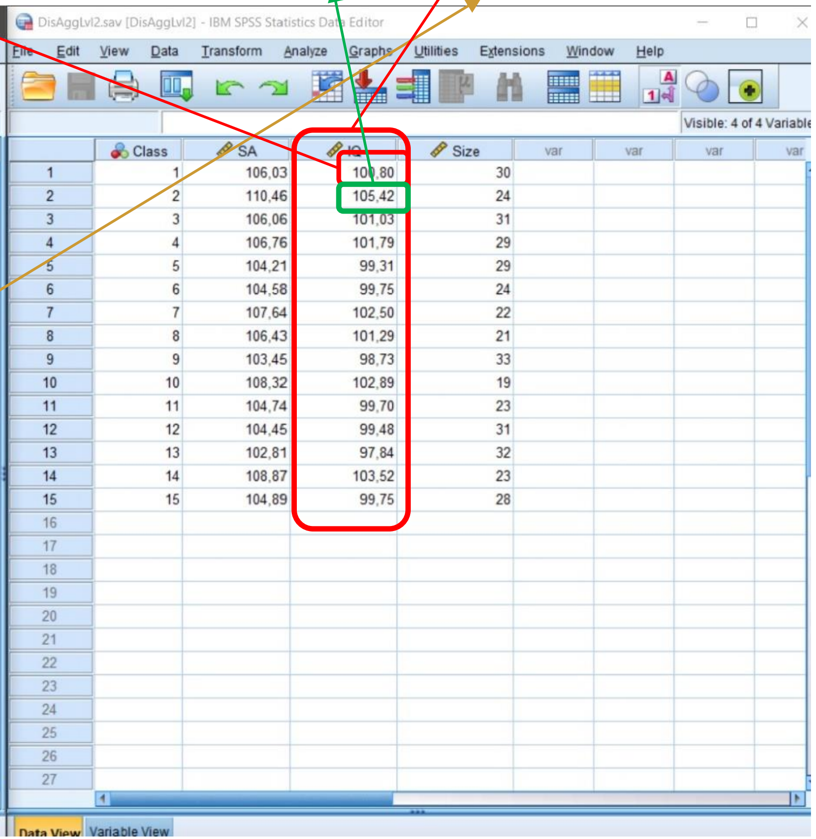
Class

Pupils





	Class	Student	Gender	SA	IQ	Size	var	var
1	1,00	1	1	106	100	30		
2	1,00	2	1	94	88	30		
3	1,00	3	2	107	102	30		
4	1,00	4	1	121	116	30		
5	1,00	5	1	109	104	30		
6	1,00	6	2	100	95	30		
7	1,00	7	1	101	96	30		
8	1,00	8	1	95	89	30		
9	1,00	9	1	98	92	30		
10	1,00	10	1	114	110	30		
11	1,00	11	1	121	115	30		
12	1,00	12	1	107	101	30		
13	1,00	13	2	97	93	30		
14	1,00	14	1	111	107	30		
15	1,00	15	2	100	96	30		
16	1,00	16	2	91	87	30		
17	1,00	17	2	110	104	30		
18	1,00	18	1	97	90	30		
19	1,00	19	2	111	106	30		
20	1,00	20	1	107	102	30		
21	1,00	21	1	117	112	30		
22	1,00	22	2	104	99	30		
23	1,00	23	1	117	110	30		
24	1,00	24	1	109	105	30		
25	1,00	25	2	99	93	30		
26	1,00	26	2	112	107	30		
27	1,00	27	1	113	108	30		



	Class	SA	IQ	Size	var	var	var	var
1	1	106,03	100,80	30				
2	2	110,46	105,42	24				
3	3	106,06	101,03	31				
4	4	106,76	101,79	29				
5	5	104,21	99,31	29				
6	6	104,58	99,75	24				
7	7	107,64	102,50	22				
8	8	106,43	101,29	21				
9	9	103,45	98,73	33				
10	10	108,32	102,89	19				
11	11	104,74	99,70	23				
12	12	104,45	99,48	31				
13	13	102,81	97,84	32				
14	14	108,87	103,52	23				
15	15	104,89	99,75	28				
16								
17								
18								
19								
20								
21								
22								
23								
24								
25								
26								
27								

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
IQ	399	70	126	100,72	9,196
Valid N (listwise)	399				

$$IQ_c = IQ_{ij} - 100.72$$

$$IQ_{c\_group} = IQ_{ij} - \overline{IQ}_j$$

Centering

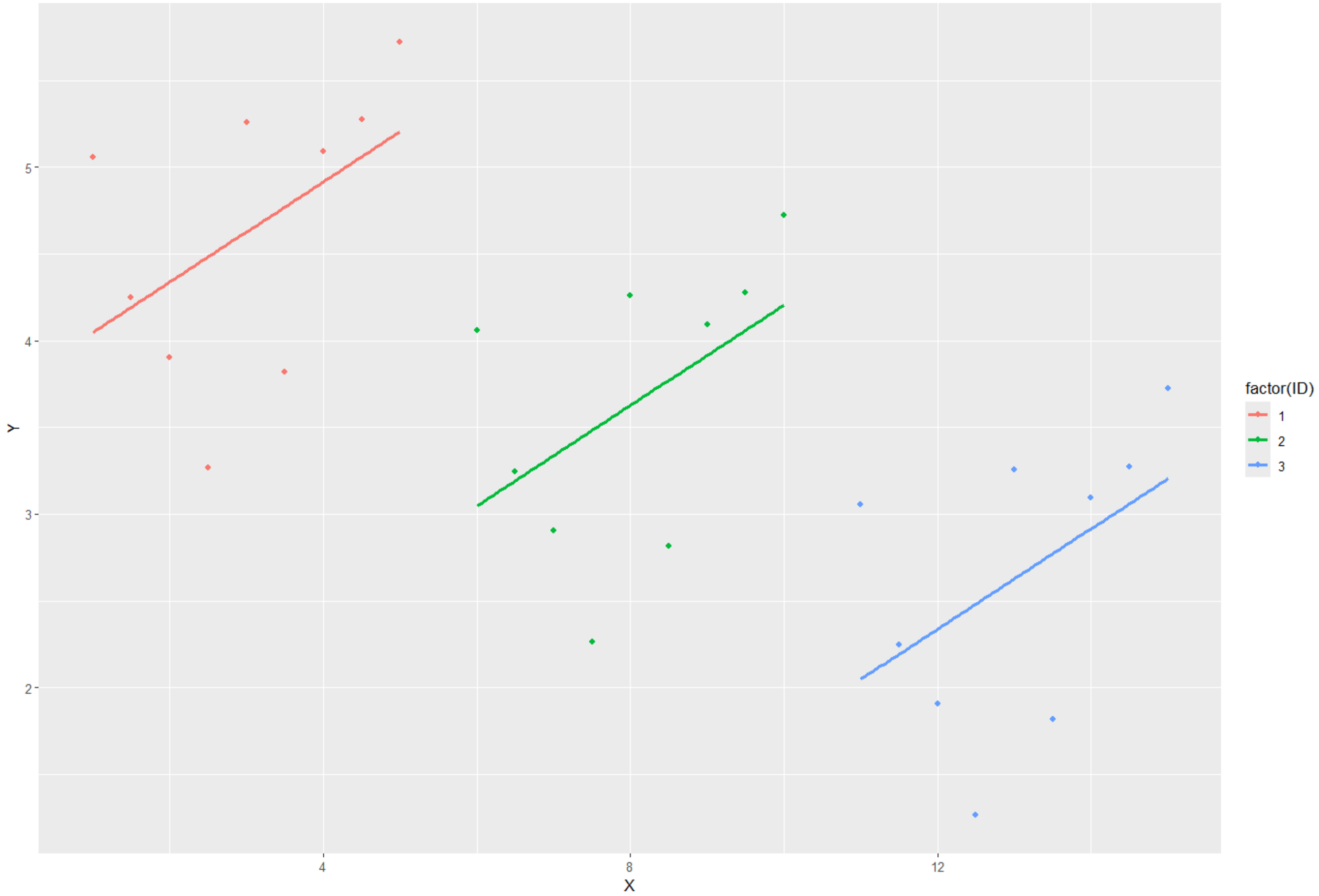
- Group-mean centering changes interpretation of variable:
  - No longer "raw" score.
  - Relative position within higher level unit.

- Used for so-called frog-pond effects:
  - Someone's actual level of an attribute doesn't matter, what matters is how one compares to the rest of their group (or pond).

- Why use this method of centering if it complicates model?

Centering

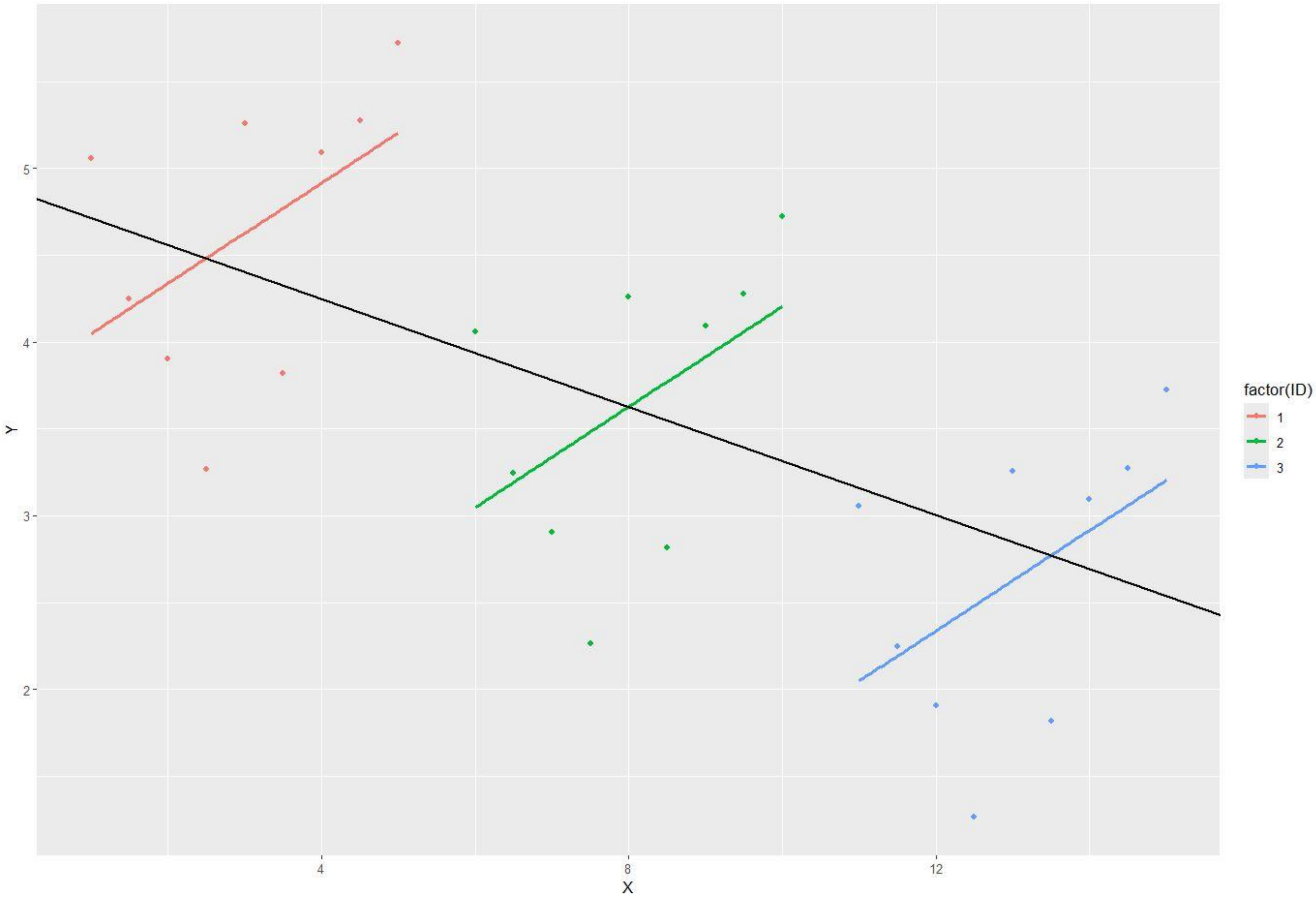
- Consider the following data of three different individuals.
- Same (positive) relation between X and Y for all three.
- Different mean scores on X and Y.



TILBURG UNIVERSITY
tesc

Centering

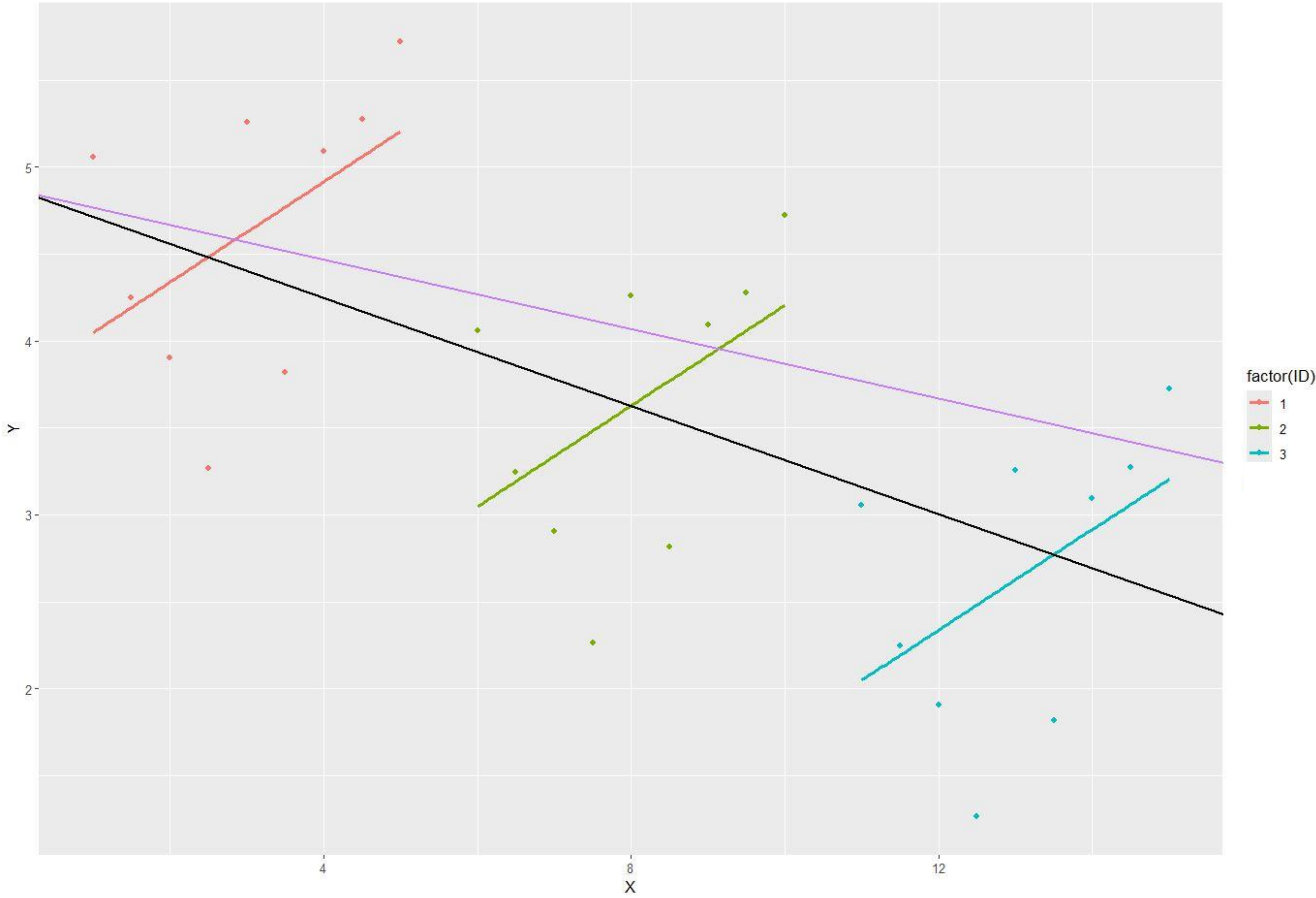
- Because the participants with the higher mean Y score have the lower mean X scores, the overall effect is negative.



Centering

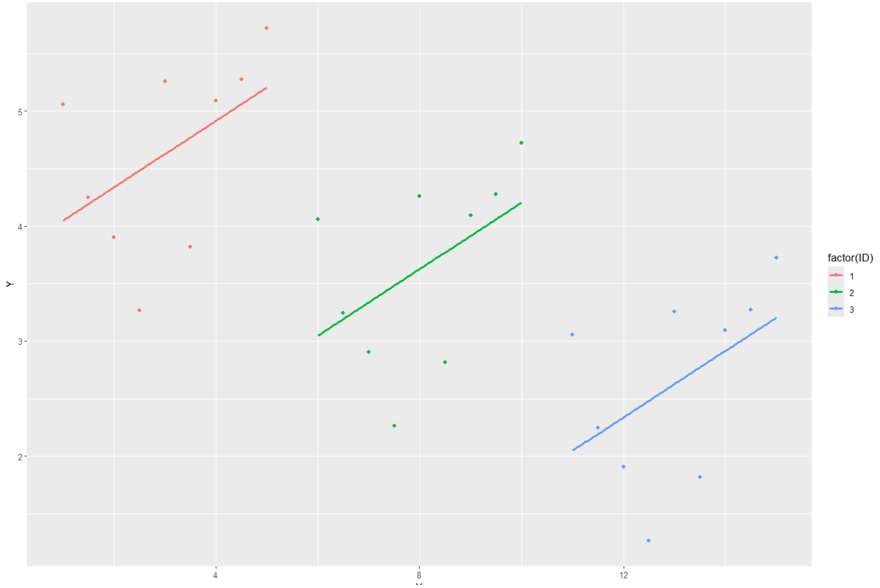
- Because the participants with the higher mean Y score have the lower mean X scores, the overall effect is negative.

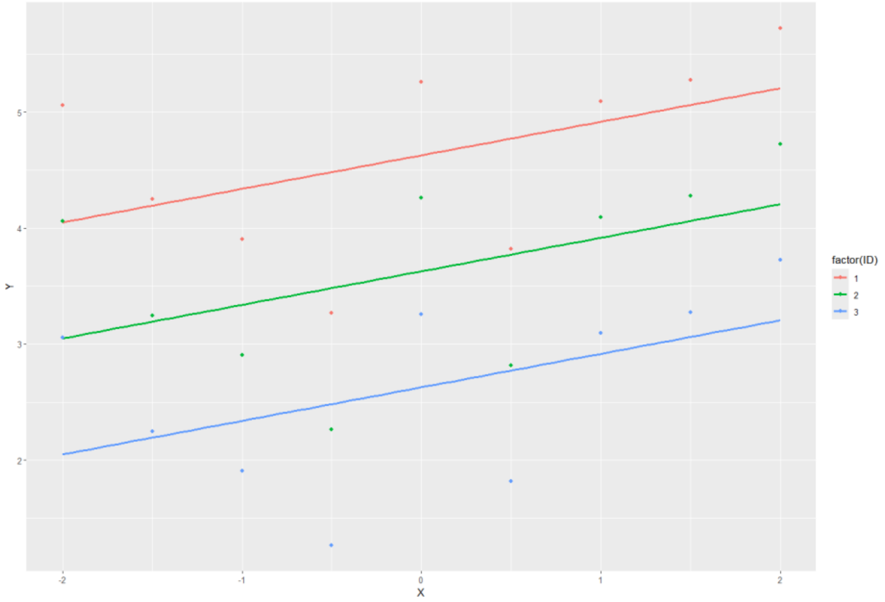
- Without group-mean centering, or fixed effect will be an uninterpretable mix of the negative overall effect and the positive within-effects



Centering

- Group-mean centering removes the between person differences on X.

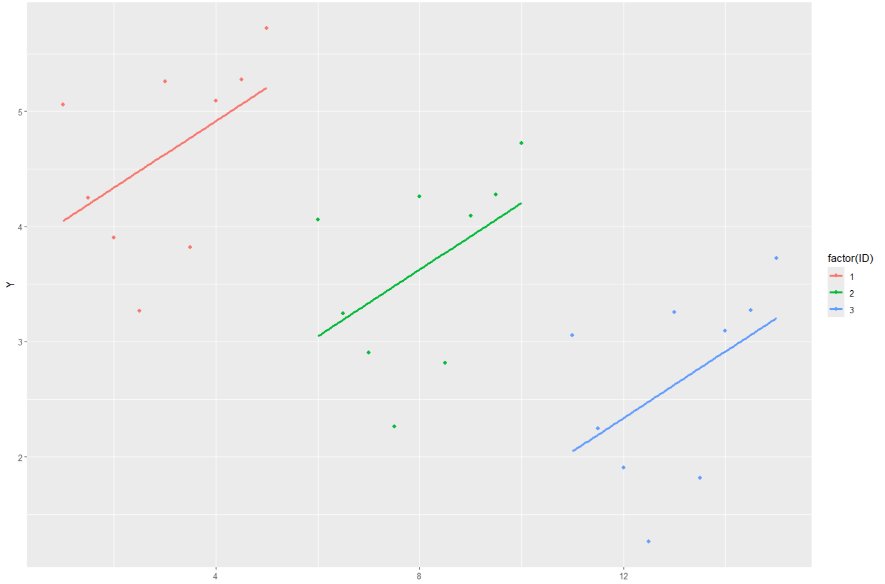


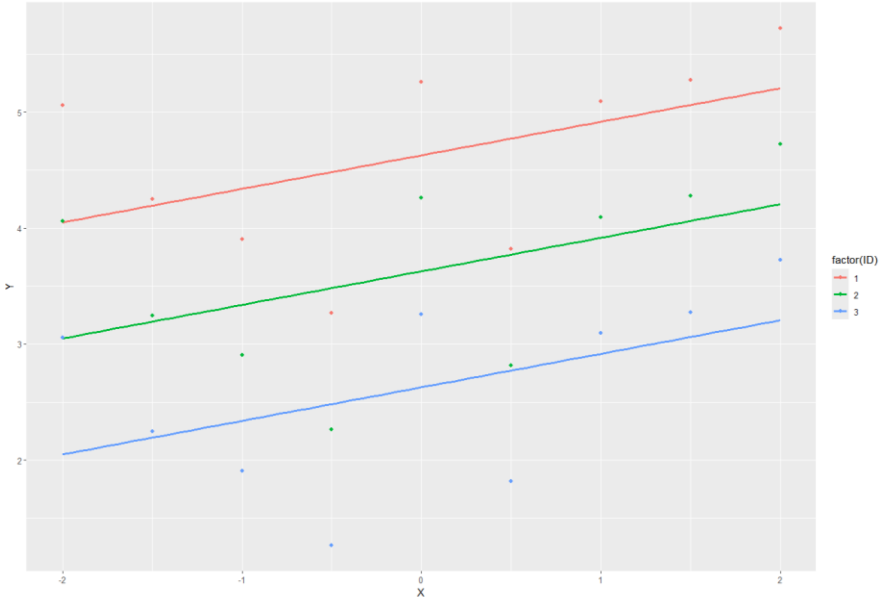


Centering

- Group-mean centering removes the between person differences on X.

- This makes the overall effect equal to the within-effects

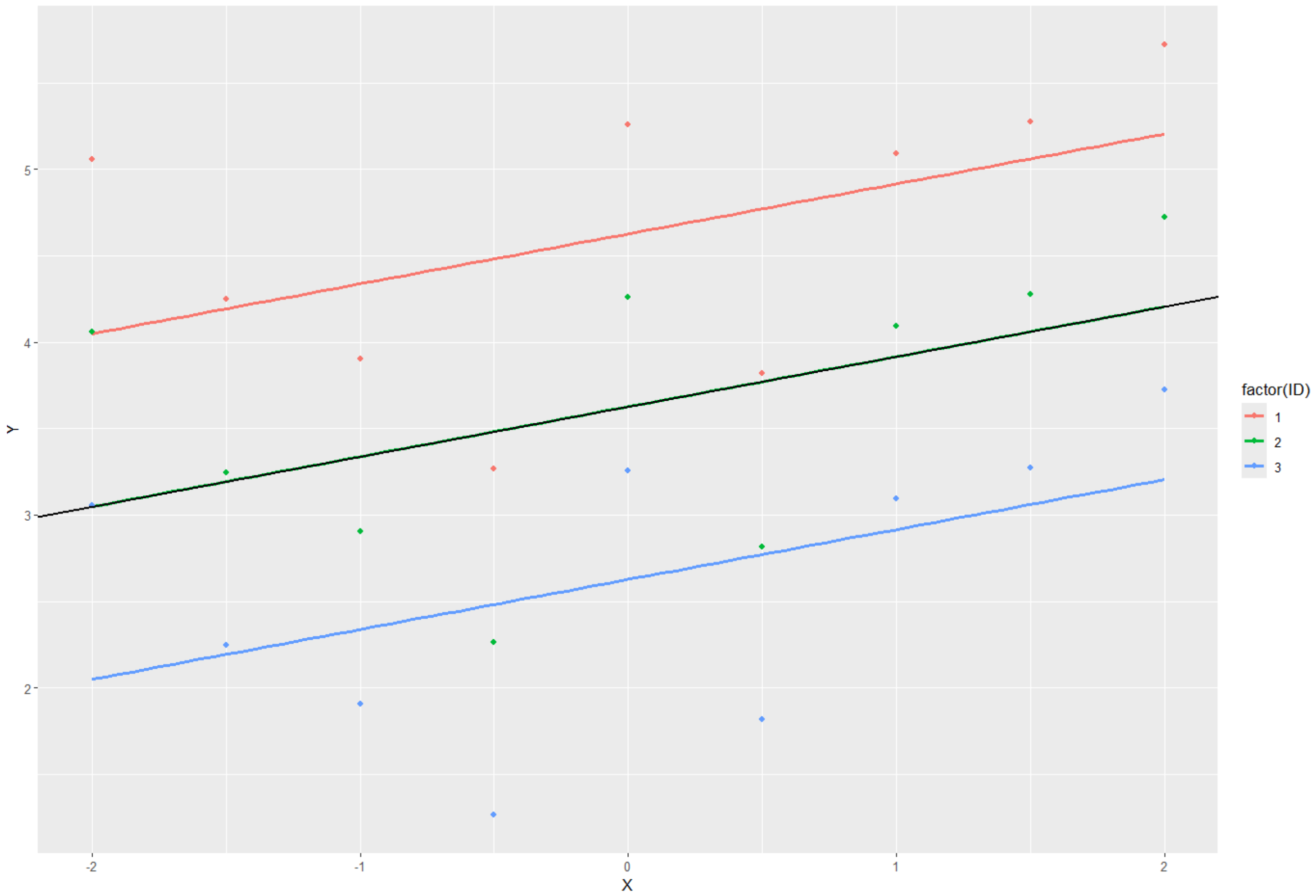




Centering

- Notice the black line for the overall effect in the middle.

- It matches the individual lines for each participant.



TILBURG UNIVERSITY

tesc

Centering

- With random slopes important to think about the intercept, and therefore the meaning of a 0-score on predictors.

- Centering advised, but…2 methods:
  - Grand-mean centering and
  - Group-mean centering.

- Group-mean centering “complicates” interpretation of model/predictor, but also gives pure estimate of within-effect.
  - So prefered if level 1 effect (or cross-level interaction) is of main interest.

Methodological Considerations: Small Level 2 N

Small Level 2 N

- Remember that with multilevel analyses, we are also modeling on level 2.
  - You wouldn't run a regression on a sample size of N = 4.

Level 1 :

$$popularity_{ij} = b_{intercept\_j} + b_{ext\_j} extraversion_{ij} + e_{ij}$$

Level 2:

$$\boldsymbol{b_{intercept\_j}} = \gamma_{intercept\_intercept} + \gamma_{texp\_intercept} * \boldsymbol{texp_j} + \boldsymbol{u_{intercept\_j}}$$

$$\boldsymbol{b_{ext\_j}} = \gamma_{intercept\_ext} + \gamma_{texp\_ext} * \boldsymbol{texp_j} + \boldsymbol{u_{ext\_j}}$$

Small Level 2 N

- Remember that with multilevel analyses, we are also modeling on level 2.
  - You wouldn't run a regression on a sample size of N = 4.
- Moreover, we are estimating variances on level 2.
  - Variances need even more observations to estimate accurately than regression coefficients or means.

Level 1 :

$$popularity_{ij} = b_{intercept\_j} + b_{ext\_j}extraversion_{ij} + e_{ij}$$

Level 2:

$$\boldsymbol{b_{intercept\_j} = \gamma_{intercept\_intercept} + \gamma_{texp\_intercept} * texp_j + u_{intercept\_j}}$$

$$\boldsymbol{b_{ext\_j} = \gamma_{intercept\_ext} + \gamma_{texp\_ext} * texp_j + u_{ext\_j}}$$

TILBURG UNIVERSITY

t e s c

Small Level 2 N

- Same applies to multilevel!
- If N on level 2 is small (less than $\pm$ 10), modeling variances/distributions on that level is probably not a great idea.
- Can you think of another way to correct for dependency in a model? To include level 2 group-membership?

Small Level 2 N

- The solution is not cluster robust SE as is sometimes suggested
  - The corrections for the SEs also need to be estimated afterall.

- How have you added groups to a regression analysis in the past?

Small Level 2 N

- The solution is not cluster robust SE as is sometimes suggested
  - The corrections for the SEs also need to be estimated afterall.
- How have you added groups to a regression analysis in the past?
  - Dummy variables!!
- With small level 2 N, it's better to account for level 2 group membership using a form of Dummy variables.

Small Level 2 N

- We're going to add a Dummy variable for every level 2 unit AND remove the intercept.

- If we have 3 classes in which we measures Popularity and Extraversion:

$$popularity_{ij} = b_{D1} * D_{Class\_1} + b_{D2} * D_{Class\_2} + b_{D3} * D_{Class\_3} + e_{ij}$$

- Can also just add level 1 predictors:

$$popularity_{ij} = b_{D1} * D_{Class_1} + b_{D2} * D_{Class_2} + b_{D3} * D_{Class_3} + b_{ext} * Extraversion_{ij} + e_{ij}$$

Small Level 2 N

$$popularity_{ij} = \ b_{D1} * D_{Class_1} + b_{D2} * D_{Class_2} + b_{D3} * D_{Class_3} + b_{ext} * Extraversion_{ij} + e_{ij}$$

- Called a Fixed Effects Model and used extensively in Economics.
- The Dummies account for all differences between the level 2 units
  - Also unmodeled level 2 influences so very robust against missed predictors.
- Level 1 effects unbiased.

Small Level 2 N

$$popularity_{ij} = \; b_{D1} * D_{Class_1} + b_{D2} * D_{Class_2} + b_{D3} * D_{Class_3} + b_{ext} * Extraversion_{ij} + e_{ij}$$

- But no free lunch!

- You can’t model level 2 predictors.
  - Because all level 2 variance is already captured by the dummies they are perfectly colinear with level 2 predictors.
  - Not a big problem, if level 2 N is small what do you hope to find there anyway?

- Can model interactions between level 1 and level 2 predictors (but requires many additional terms, you need to add interactions for all Dummies).