

# Multilevelanalyse

Workshop 28 januari 2015

Peter Verboon, Denise Peels, Ron Pat-EI

# Inleiding

- Wat zijn multilevel data
- Fixed en Random effecten
- Multilevel Modellen
- Uitvoeren van een multilevel analyse
- Interpretieren van de uitkomsten van multilevel analyse

## Wat zijn multilevel data?

Data zijn niet onafhankelijk

Data zijn gegroepeerd binnen niveaus' of levels, bv 2 levels:

Leerlingen → Klassen

Werknemers → Bedrijven

Meetmomenten → Personen

Voorbeelden 3 levels:

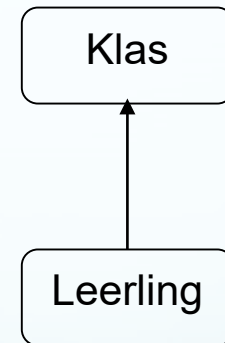
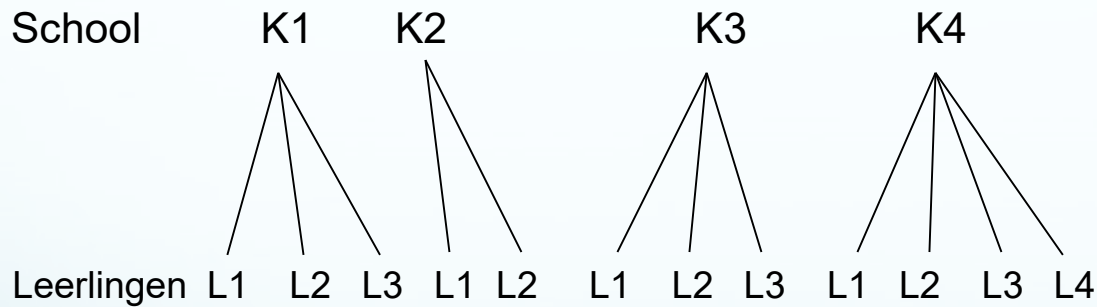
Leerlingen → Klassen → Scholen

Werknemers → Bedrijven → Sectoren

Meetmomenten → Dagen → Personen

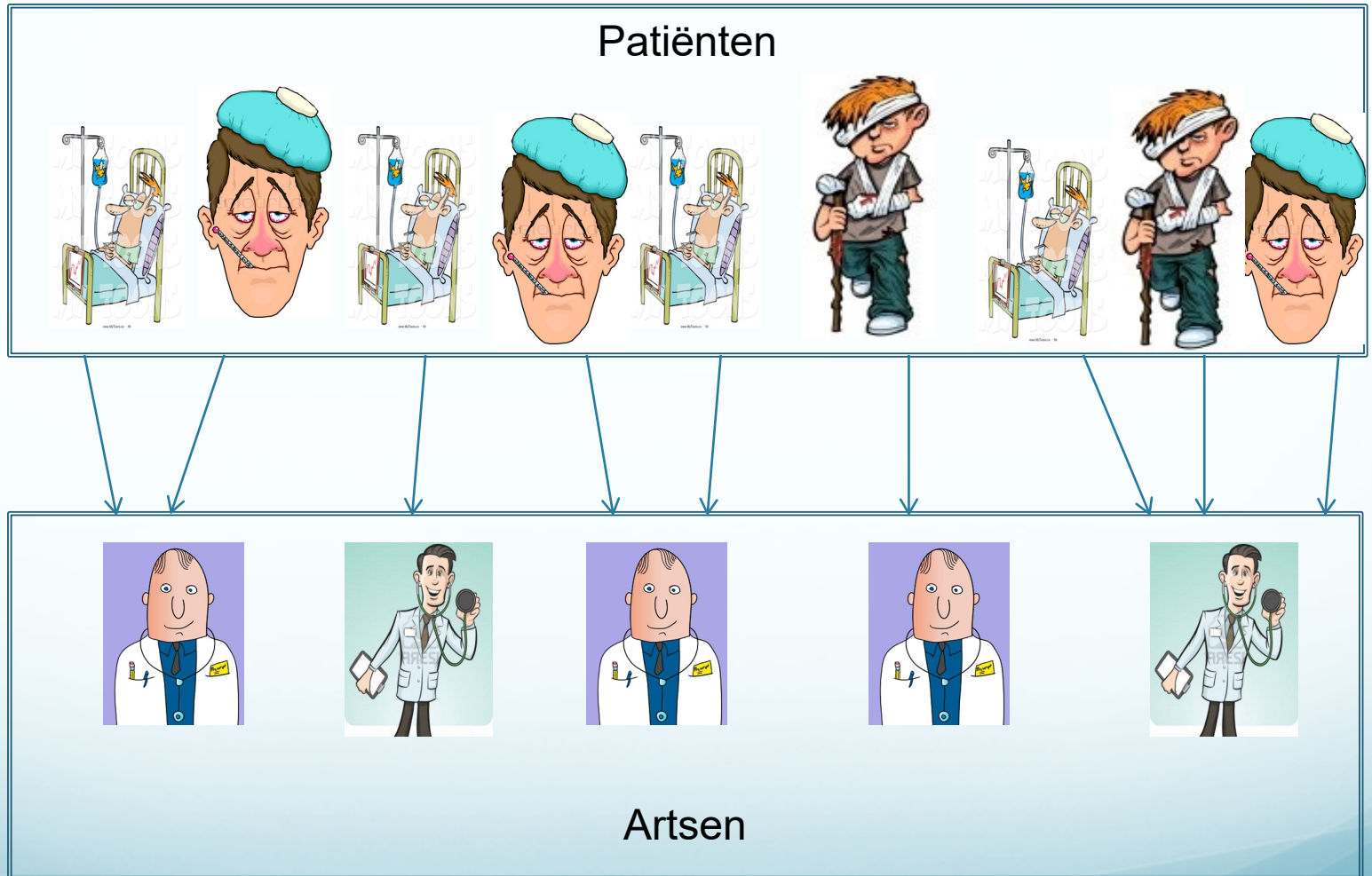
# Hierarchische structuur

Leerlingen binnen klassen

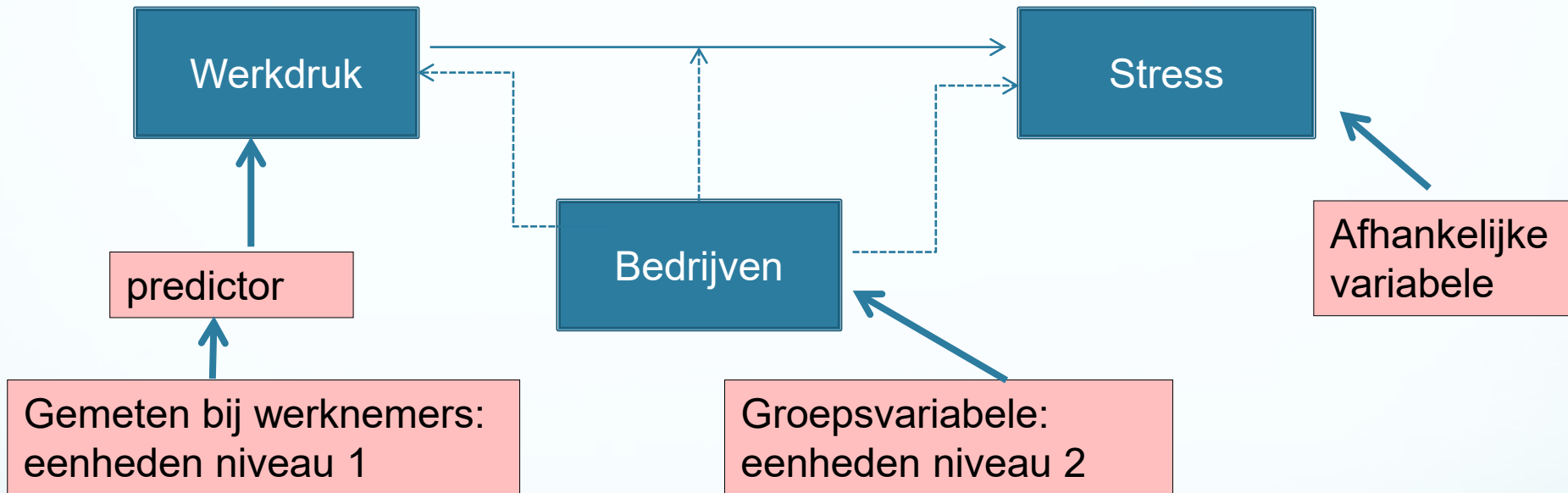


Leerlingen binnen een klas lijken meer op elkaar dan een random steekproef van leerlingen → cluster effect

# Schematische weergave



# Voorbeeld: Conceptueel model

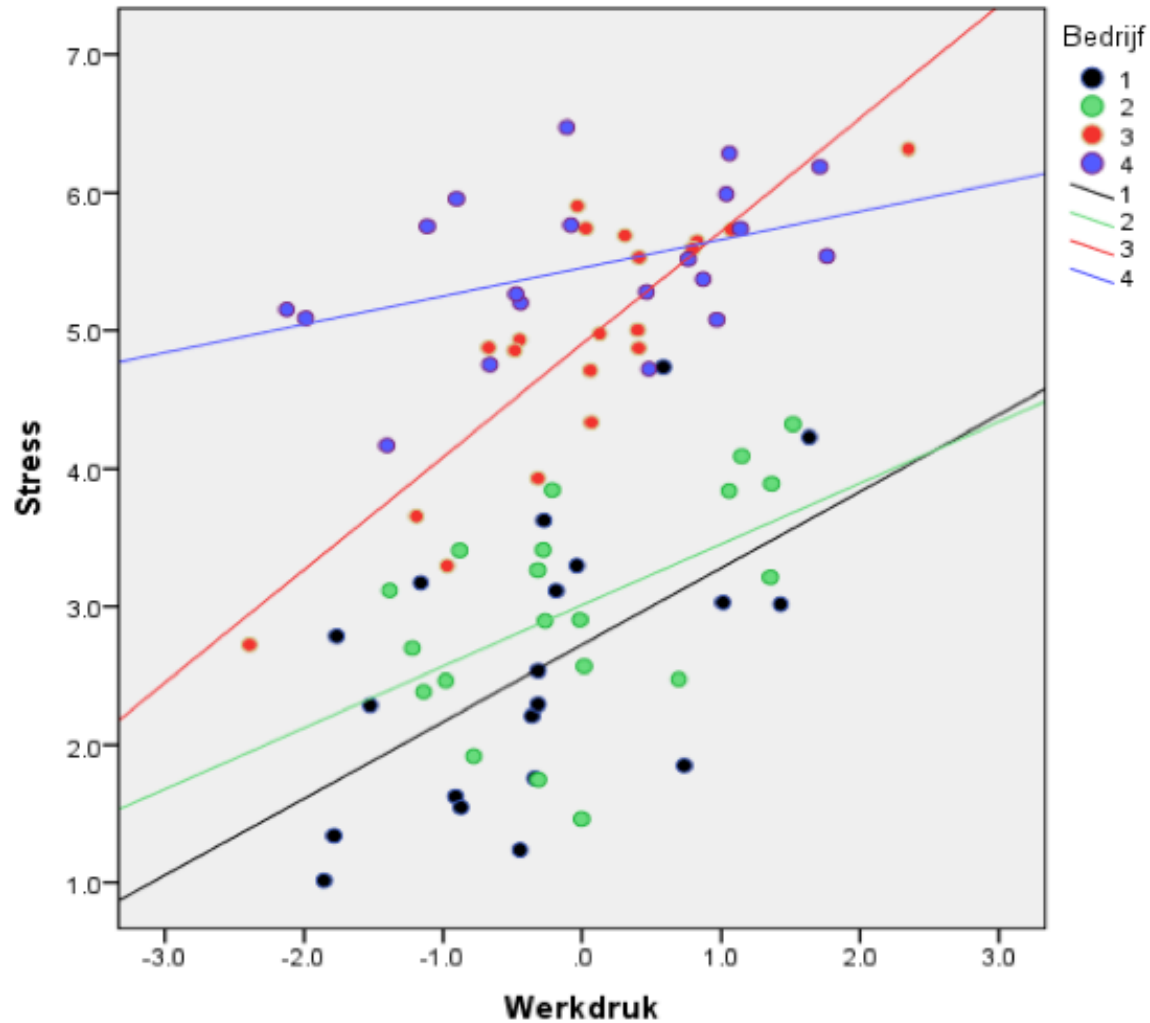


Hypothese: Werkdruk veroorzaakt na verloop van tijd Stress

Maar:

- (1) Hoogte van de Stress en/of Werkdruk verschillen mogelijk per bedrijf
- (2) Relatie tussen Werkdruk en Stress verschilt mogelijk per bedrijf

# Voorbeeld: verband werkdruk en stress bij 4 bedrijven



## Fixed en Random effecten

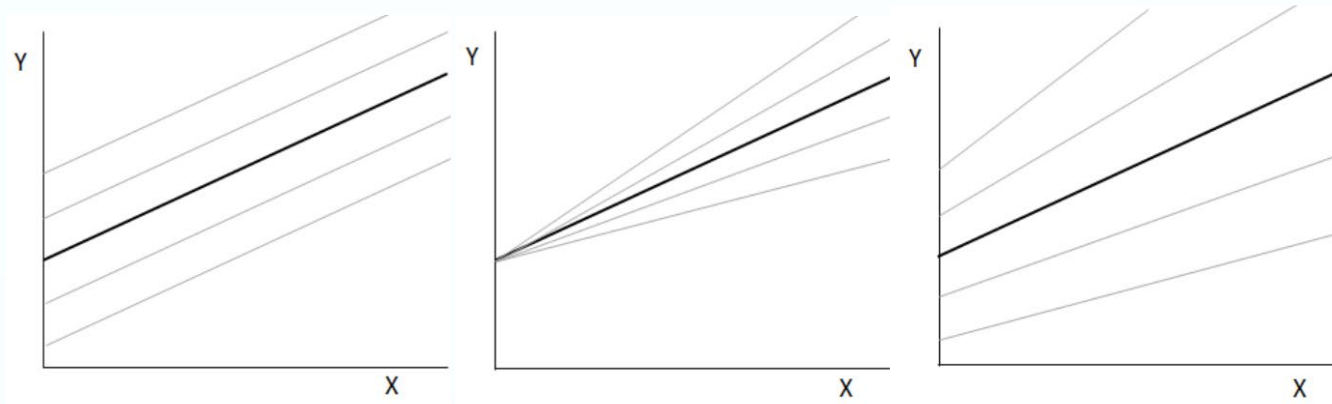
- Fixed effecten: effecten die **niet** variëren over de groepen
- Random effecten: effecten die **wel** variëren over de groepen
- Twee type effecten:
  - Intercepten, slopes

Slope = richtingscoëfficiënt van een lijn  
= regressie-coëfficiënt

Intercept = snijpunt lijn met y-as  
= gemiddelde van de AV als alle  
predictoren zijn gecentreerd



## Random versus Fixed intercept en slope



- a. Random intercept, fixed slope
- b. Fixed intercept en random slope
- c. Random intercept en random slope

# Regressie versus Multilevel Analyse

- **Regressie Analyse**

Geen random effecten verondersteld

Eventueel aanwezige groepen liggen vast (“fixed”), zijn niet vervangbaar door andere in de populatie

Alleen niveau 1 bevat random elementen

- **Multilevel Analyse**

Wel random effecten verondersteld

Groepen vormen een random steekproef uit populatie met groepen

Meerdere niveaus bevatten random elementen

## Soorten Multilevel Data

- Hiërarchische structuur

Een afhankelijke variabele meten bij respondenten binnen groepen

- Herhaalde metingen

Dezelfde variabelen herhaaldelijk meten in de tijd bij respondenten

# Multilevelmodel

Standaard regressiemodel:

$$Y_i = b_0 + b_1 X_i + e_i,$$

$i$  = index van b.v. kinderen

Basis multilevelmodel:

$$Y_{ij} = b_{0j} + b_{1j} X_{ij} + e_{ij},$$

$j$  = index van b.v. klassen

## Intercept-Only model (IO-model)

Het intercept (en verder niets) verschilt per klas:

**Niveau 1:**  $WIS = b_{0j} + e_{ij}$ .

$i$  = index van kinderen

$j$  = index van klassen

**Niveau 2:**  $b_{0j} = \gamma_{00} + u_{0j}$ .

→  $WIS = \gamma_{00} + u_{0j} + e_{ij}$ .

## Intraclass correlatie

Intraclass correlatie (ICC) geeft aan de verhouding van de variantie binnen de klassen ten opzichte van de totale variantie van het criterium.

$$\text{ICC} = \frac{\text{variantie tussen groepen}}{\text{variantie tussen groepen} + \text{variantie binnen groepen}}$$

$$\text{ICC} = \frac{u_{0j}}{u_{0j} + e_{ij}}$$

# Syntax IO-model

MIXED WIS



Afhankelijke variabele

/PRINT= SOLUTION TESTCOV



Gewenste output

/METHOD= ML



Optimaliseringsmethode

/FIXED= INTERCEPT



Fixed effecten in het model

/RANDOM= INTERCEPT | SUBJECT(KLAS).



Random effecten in  
het model



Variabele die random  
effecten veroorzaakt

# Output IO-model

Tabel 4. Resultaten van de analyse van model A.

Estimates of Fixed Effects <sup>a</sup>							
Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	5.834	.242	9.820	24.079	.000	5.281	6.373

Estimates of Covariance Parameters <sup>a</sup>							
Parameter	Estimate	Std. Error	Wald Z	Sig.	95% Confidence Interval		
					Lower Bound	Upper Bound	
Residual	2.317	.240	9.639	.000	1.890	2.839	
Intercept [subject = Klas]	.468	.265	1.767	.077	.154	1.421	

↑  
**Varianties**

Conclusie: Het gemiddelde wiskundecijfer is 5.8, maar dit gemiddelde verschilt (bijna significant) per klas



# Output IO-model

Tabel 4. Resultaten van de analyse van model A.

Estimates of Fixed Effects <sup>a</sup>							
Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	5.834	.242	9.820	24.079	.000	5.281	6.373

Estimates of Covariance Parameters <sup>a</sup>						
Parameter	Estimate	Std. Error	Wald Z	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Residual	2.317	.240	9.639	.000	1.890	2.839
Intercept [subject = Klas]	.468	.265	1.767	.077	.154	1.421

↑  
**Varianties**

Intraclass correlatie:  $0.47 / (0.47 + 2.32) = .168 \rightarrow 17\%$

## Past het model bij de data (Fit)

Maat voor passendheid of fit:

-2 x Loglikelihood (-2LL)

Geen *absoluut* criterium, maar bij **geneste** modellen wel een *relatief* criterium.

Wordt ook wel met term “Deviance” aangeduid

## Centreren van variabelen

Twee mogelijkheden:

Centreren tov de totale dataset (totaal van variabele is nul)

Centreren binnen de groepen (binnen iedere groep heeft variabele gemiddelde nul)

Bij groepscentreren is het mogelijk en soms nuttig om de groepsgemiddelden als extra variabele mee te nemen in het model.

## Model met 1 predictor

IQ voorspelt Wiskundescores : random intercept, fixed slope

$$\text{WIS} = b_{0j} + b_1 \text{IQ} + e_{ij}$$

fixed random fixed

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

## Model met 1 predictor

IQ voorspelt Wiskundescores : random intercept, random slope

$$\text{WIS} = b_{0j} + b_{1j} \text{IQ} + e_{ij}$$

fixed      random

$$b_{0j} = \gamma_{00} + u_{0j}$$
$$b_{1j} = \gamma_{10} + u_{1j}$$

$\gamma_{10}$  fixed  
 $u_{1j}$  random

# Resultaten MLA

Tabel 6. Resultaten van de analyse van model C.

*Estimates of Fixed Effects<sup>a</sup>*

Parameter	Estimate	Std.	df	t	Sig.	95% Confidence Interval	
		Error				Lower Bound	Upper Bound
Intercept	5.870	0.170	10.088	34.517	.000	5.492	6.249
IQ	1.149	0.187	10.152	6.134	.000	0.732	1.566

*Estimates of Covariance Parameters<sup>a</sup>*

Parameter		Estimate	Std.	Wald Z	Sig.	95% Confidence Interval	
			Error			Lower Bound	Upper Bound
Residual		0.693	0.073	9.392	.000	0.562	0.854
Intercept [subject = Klas]	Var	0.250	0.128	1.946	.052	0.091	0.685
IQ [subject = Klas]	Var	0.311	0.155	2.003	.045	0.117	0.829

a. Dependent Variable: Wiskundescore.

Totaal 5 effecten: 2 fixed en 3 random

# Opdracht 1

## Opdracht 1 (ca. 30 minuten)

**Eenvoudige multilevel modellen met een continue uitkomstmaat**

## Model met interactie

IQ voorspelt Wiskundescores : random intercept, random slope

$$\text{WIS} = b_{0j} + b_{1j} \text{IQ} + e_{ij}$$

$\gamma_{10}$  fixed

$u_{1j}$  random

fixed

random

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

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

Hier kunnen predictoren (moderatoren)  
worden toegevoegd



## Model met interactie

IQ voorspelt Wiskundescores met extra predictoren : random intercept, random slope

$$WIS = b_{0j} + b_{1j} IQ + e_{ij}$$

$$b_{0j} = \gamma_{00} + u_{0j} + \gamma_{01} ERV$$

$$b_{1j} = \gamma_{10} + u_{1j} + \gamma_{11} Aantal$$

$$WIS = (\gamma_{00} + u_{0j} + \gamma_{01} ERV) + (\gamma_{10} + u_{1j} + \gamma_{11} Aantal) * IQ + e_{ij}$$



**Interactieterm:**  $\gamma_{11} Aantal * IQ$

## Syntax voor interactie

```
MIXED WIS WITH IQ Aantalc ERVc  
/PRINT=SOLUTION TESTCOV  
/METHOD= REML  
/FIXED= INTERCEPT IQ ERVc Aantalc IQ*Aantalc  
/RANDOM= INTERCEPT IQ | SUBJECT(Klas).
```

Achter "FIXED=" staat een gewoon regressie-model

## Opdracht 2

### Opdracht 2 (ca. 20 minuten)

### **Multilevel modellen met een interactie-effect**

# Herhaalde metingen

Veel voorkomende toepassing van MLA

Meerdere metingen bij een enkel persoon

De metingen of tijdstippen vormen het eerste niveau

Personen vormen het tweede niveau

Samenhang tussen metingen ontstaat vanwege:

1. variatie binnen personen  $<$  variatie tussen personen
2. metingen (kort op elkaar in de tijd) hangen vaak samen

## Dagboekmethode

Dagboekmethode of variaties daarop: mensen vullen een aantal keer per dag een aantal gegevens in. Vaak meerdere dagen achter elkaar.

B.v. stressniveau en de zin in een sigaret

Hypothese: stress leidt tot zin in een sigaret

# Syntax

MIXED ZIN WITH STRESS<sub>c</sub> METING

/PRINT=SOLUTION TESTCOV

/FIXED=INTERCEPT STRESS<sub>c</sub>

/RANDOM=INTERCEPT STRESS<sub>c</sub> | SUBJECT(RESPONDENT)

**/REPEATED=METING | SUBJECT(RESPONDENT) COVTYPE(AR1).**

Variabele die de herhaalde meting aangeeft.

Structuur van de errorterm:  
autoregressie

Alleen laatste regel is nieuw t.o.v. van de eerdere analyses.

## Herhaalde metingen

Output fixed effects zijn gelijk aan eerdere analyses.

Output van de random effects:

Estimates of Covariance Parameters<sup>a</sup>

Parameter		Estimate	Std. Error	Wald Z	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Repeated Measures	AR1 diagonal	1.723	.217	7.939	.000	1.346	2.206
	AR1 rho	.485	.065	7.408	.000	.347	.603
Intercept [subject = RESPONDENT]	Variance	.560	.283	1.979	.048	.208	1.509
STRESSc [subject = RESPONDENT]	Variance	.106	.051	2.064	.039	.041	.276

a. Dependent Variable: ZIN.

Vrij hoge correlatie tussen de opeenvolgende metingen

## Opdracht 3

### **Opdracht 3 (ca. 30 minuten)**

### **Multilevel analyse van herhaalde metingen**



## MLA met een dichotome uitkomst

MLA kan ook als uitkomstmaat dichotoom is.

Zo'n uitkomstmaat komt regelmatig voor:

Bepaald gedrag wordt wel of niet vertoond (b.v. roken).

Iemand wordt wel of niet aangenomen of ontslagen.

Wat is er anders ten opzichte van eerdere analyses?

## Dichotome uitkomst

Bij dichotome uitkomst wordt meestal logistische regressie i.p.v. lineaire regressie toegepast

De interpretatie van de regressie-coëfficiënten is lastiger bij logistische regressie

SPSS gebruikt wel een andere procedure, namelijk GENLINMIXED i.p.v. MIXED

Maar voor *het principe* van de Multilevel aanpak verandert er verder niet zo veel

## Logistische regressie

$$\ln\left(\frac{p}{1-p}\right) = y^* = a + b_1x_1 + b_2x_2 + r$$

$p$  = kans op bepaalde uitkomst

$$\frac{p}{1-p} = \text{odds} = e^{a+bx+r}$$

$$p = \text{odds} / (1 + \text{odds})$$

# Logistische regressie

Interpretatie van de coëfficiënten

$$\frac{\text{odds}(x+1)}{\text{odds}(x)} = \frac{e^{a+b(x+1)}}{e^{a+bx}} = e^{a+b(x+1)-a-bx} = e^b$$

$$e^b = \text{odds-ratio}$$

B.v. Als  $e^b = 2$  dan neemt de kansverhouding (odds) een factor 2 toe, als je variabele  $x$  met 1 verhoogt

# Syntax MLA met Dichotome uitkomst

## GENLINMIXED

```
/DATA_STRUCTURE SUBJECTS= RESPONDENT
```

```
    REPEATED_MEASURES= METING COVARIANCE_TYPE= AR1
```

```
/FIELDS TARGET= ROKEN
```

```
/TARGET_OPTIONS DISTRIBUTION= BINOMIAL LINK= LOGIT
```

```
/FIXED EFFECTS= STRESSc USE_INTERCEPT= TRUE
```

```
/RANDOM EFFECTS= STRESSc USE_INTERCEPT= TRUE
```

```
    SUBJECTS= RESPONDENT COVARIANCE_TYPE=
```

```
    VARIANCE_COMPONENTS
```

```
/BUILD_OPTIONS TARGET_CATEGORY_ORDER= DESCENDING.
```

## Output MLA met Dichotome uitkomst

Fixed Coefficients						
Target: ROKEN , Reference Category: 0						
Parameter	Estimate	Std. Error	t	Sig.	95% Confidence Interval	
					Lower	Upper
Intercept	-1.609	.289	-5.561	.000	-2.178	-1.039
STRESSc	0.913	.247	3.691	.000	0.426	1.400

Interpretatie makkelijker in termen van odds (ipv log-odds):

$$e^{0.913} = \exp(0.913) = 2.49$$

Als stress een eenheid toeneemt, dan neemt de kansverhouding (odds van roken ten opzichte van niet roken) met circa 2.5 toe.

## Output MLA met Dichotome uitkomst

Fixed Coefficients						
Target: ROKEN , Reference Category: 0						
Parameter	Estimate	Std. Error	t	Sig.	95% Confidence Interval	
					Lower	Upper
Intercept	-1.609	.289	-5.561	.000	-2.178	-1.039
STRESSc	0.913	.247	3.691	.000	0.426	1.400

Interpretatie makkelijker in termen van odds (ipv log-odds):

$$e^{0.913} = \exp(0.913) = 2.49$$

Als stress een eenheid toeneemt, dan neemt de kansverhouding (odds van roken ten opzichte van niet roken) met circa 2.5 toe.

## Output MLA met Dichotome uitkomst

### Estimates of Covariance Parameters

Parameter		Estimate	Std. Error	95% Confidence Interval	
				Lower	Upper
Repeated Measures	AR1 diagonal	.657	.060	0.549	0.786
	AR1 rho	-.050	.070	-0.185	0.087
Intercept [subject = RESPONDENT]	Variance	1.181	.630	0.415	3.358
STRESSc [subject = RESPONDENT]	Variance	.780	.433	0.263	2.315

De autocorrelatie tussen de meetmomenten is hier dus gering (-.050) en niet significant.



## Output MLA met Dichotome uitkomst

-2LL =	1493.52
Predicted Correct	80.3%
ROKEN = 0	93.8%
ROKEN= 1	40.0%

Het model kan de niet-rokers beter voorspellen omdat er daar meer van zijn.

# Opdracht 4

## Opdracht 4 (ca. 30 minuten)

### **Multilevel analyse met een dichotome uitkomst**

## MLA met meerdere levels

Data kunnen zijn genest in meer dan drie niveau's

Voorbeelden 3 levels:

Leerlingen → Klassen → Scholen

Werknemers → Bedrijven → Sectoren

Meetmomenten → Dagen → Personen

Voeg in SPSS-Syntax eventueel een extra random statement toe.

## MLA met meerdere levels

Voorbeeld data: Stoppen\_met\_Roken

Model: Stress + Pos. Expectations → intentie tot stoppen

ESM methode: 10 metingen per dag per subject gedurende 7 dagen

2 levels: metingen genest binnen dagen genest binnen subjecten

Indien geen missing:  $7 \times 10 = 70$  metingen per subject

## MLA met meerdere levels

Model:  $Y_{idj} = b_{0dj} + b_{1dj}X_{idj} + e_{idj}$ ,

Subject  $i$  ( $i=1, \dots, N$ );

dag  $d$  ( $d=1, \dots, 7$ );

meetmoment  $j$  ( $j=1, \dots, 10$ )

Mogelijke aanname:  $b$  (intercept of slope) hangt af de dag van meting of van het subject of van het subject op een bepaalde dag.

Meetmomenten (residuen) zijn mogelijk onderling afhankelijk → repeated measures met AR1 structuur

## SPSS Syntax bij MLA met 3 levels

Model met random intercept mbt tot subjecten en mbt tot dagen (binnen subjecten)

MIXED intentie WITH Stressc POEsc

/PRINT= SOLUTION TESTCOV

/METHOD=REML

/FIXED=Stressc POEsc | SSTYPE(3)

**/RANDOM=INTERCEPT | SUBJECT(subjnr) COVTYPE(VC)**

**/RANDOM=INTERCEPT | SUBJECT(subjnr\*dagnr) COVTYPE(VC)**

/REPEATED=beepnr | SUBJECT(subjnr\*dagnr) COVTYPE(AR1).

# SPSS output bij MLA met 3 levels

`/REPEATED=beepnr | SUBJECT(subjnr*dagnr) COVTYPE(AR1)`

Parameter		Estimate	Std. Error	Wald Z	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Repeated Measures	AR1 diagonal	.758331	.026571	28.540	.000	.708001	.812238
	AR1 rho	.332582	.023456	14.179	.000	.285848	.377737
Intercept [subject = subjnr]	Variance	.799214	.172745	4.627	.000	.523216	1.220801
Intercept [subject = subjnr * dagnr]	Variance	.155932	.027960	5.577	.000	.109725	.221597

a. Dependent Variable: ben ik van plan om van sigaretten af te blijven.

Vergelijk met random intercept binnen alleen subjects:

Estimates of Covariance Parameters<sub>a</sub>

Parameter		Estimate	Std. Error	Wald Z	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Repeated Measures	AR1 diagonal	.900234	.028547	31.535	.000	.845986	.957960
	AR1 rho	.439105	.018428	23.828	.000	.402288	.474505
Intercept [subject = subjnr]	Variance	.806762	.171248	4.711	.000	.532186	1.223000

a. Dependent Variable: ben ik van plan om van sigaretten af te blijven.

# SPSS output bij MLA met 3 levels

Model met random intercept en slopes voor Stressc:

```
/RANDOM=INTERCEPT Stressc | SUBJECT(subjnr) COVTYPE(VC)
```

```
/RANDOM=INTERCEPT Stressc | SUBJECT(subjnr*dagnr) COVTYPE(UN)
```

Estimates of Covariance Parameters<sup>a</sup>

Parameter		Estimate	Std. Error	Wald Z	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Repeated Measures	AR1 diagonal	.743205	.026782	27.751	.000	.692525	.797594
	AR1 rho	.534374	.023893	13.994	.000	.286748	.380350
Intercept [subject = subjnr]	Variance	.813526	.175364	4.639	.000	.533196	1.241243
Stressc [subject = subjnr]	Variance	.005095	.003703	1.376	.169	.001226	.021177
Intercept + Stressc [subject = subjnr * dagnr]	UN (1,1)	-.141851	.026955	5.262	.000	.097742	.205865
	UN (2,1)	.020485	.008483	2.415	.016	.003857	.037112
	UN (2,2)	.009119	.005614	1.624	.104	.002728	.030479

a. Dependent Variable: ben ik van plan om van sigaretten af te blijven.



## Opdracht 5

### **Opdracht 5 (ca. 30 minuten)**

### **Multilevel analyse met drie levels**