Introduction à l'analyse de données quantitatives longitudinales avec les modèles de courbe de croissance

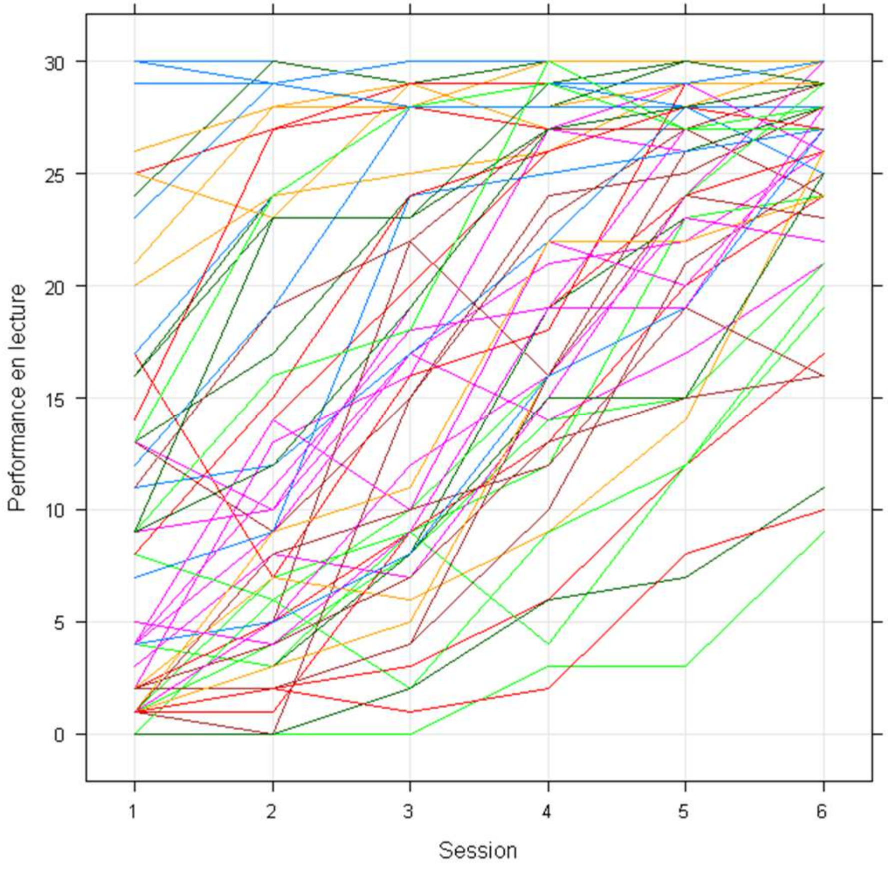
Jacques Juhel
CRPCC (E.A. 1285), Université Rennes 2

Objectifs de l'analyse de données longitudinales

- Description de trajectoires ou patterns de changement au cours du temps.
- Identification des prédicteurs des trajectoires de changement.
- Etude des relations entre trajectoires de changement et variables distales.

Par exemple

Evolution, de septembre à juin, de la performance à une tâche de lecture de mots chez 60 enfants de CP



Nombre d'occasions de mesure et possibilités de modélisation

- 3 occasions de mesure : fonction linéaire du temps (#par = 2).
- 4 occasions de mesure : fonctions polynomiale ou exponentielle du temps (#par = 3).
- 5 occasions de mesure ou plus : fonctions non linéaires du temps (#par = 4 ou plus).

Points critiques

- Mesures temporellement stables ou non?
- Occasions de mesure identiques pour tous les individus ou pas?
- Données manquantes ou pas?
- Même trajectoire pour tous les individus ou pas?

Introduction
GLM
MLR
LGC

Diversité des modèles de courbe de croissance

1. Modèle linéaire général : ANOVA (uni- et multivariée) à mesures répétées
2. a. Modèles linéaires mixtes et modèles linéaires généralisés mixtes (MLR: modèles de régression multi-niveaux)

   b. Modèles de courbe de croissance à variables latentes (LGC: famille des modèles d'équations structurelles)
3. Données hétérogènes : modèles de croissance à mélange de distributions

Burchinal, M., Nelson, L., & Poe, M. (2006). Growth curve analysis: An introduction to various methods for analyzing longitudinal data. *Monographs of the Society for Research in Child Development, 71(3),* 65-87.

1. Modèle linéaire général

Introduction
GLM
MLR
LGC

Rappels

- Distribution de probabilité d'une variable aléatoire *Y*.
- Moyenne sur la population :

$$\mu = E(Y)$$

- Variance sur la population :

$$\mathrm{var}(Y) = E\{(Y-\mu)^2\}$$

- Sources de variation :

$$Y_i = \mu + \varepsilon_i$$

$$\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$$

Introduction
GLM
MLR
LGC

Modèle des moyennes, modèle des variances

$$Y_i = \mu + \varepsilon_i$$

$$\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$$

- Composante structurale : effets fixes (test d'hypothèses)
- Composante stochastique : effets aléatoires et résidus (hypothèses de distribution)

Modèle à effets fixes (linéaire simple)

- Pour chaque valeur fixe x_i :

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0, \sigma^2).$$

- Idéalement, pour chaque valeur x_i, la variable aléatoire continue Y_i devrait être égale à sa moyenne i.e. à la valeur fixe $\beta_0 + \beta_1 x_i$.
- ε_i représente l'écart à $\beta_0 + \beta_1 x_i$ dû aux effets agrégés des variations interindividuelles et de l'erreur de mesure.

Introduction
GLM
MLR
LGC

Modèle à effets fixes (linéaire simple)

$$\varepsilon_i = \boxed{u_i} + \boxed{e_i}$$

- variations interindividuelles + erreur de mesure

$$\mathrm{var}(\varepsilon) = \sigma^2 = \sigma_u^2 + \sigma_e^2,$$
$$\mathrm{var}(Y) = \sigma^2.$$

- La modélisation statistique (≠ description des données) explicite les sources de variation.

Modèle à effets fixes (linéaire simple)

$$Y_i \sim \mathcal{N}(\beta_0 + \beta_1 x_i, \sigma^2)$$

- A chaque x_i correspond une distribution de valeurs possibles Y_i dont la moyenne est $\beta_0 + \beta_1 x_i$ et la variance σ^2.
- Hypothèse : les variables aléatoires Y_i (ou de manière équivalente les ε_i) sont indépendantes.
- Hypothèse non réaliste avec des données longitudinales!

Données répétées et longitudinales

- Soit la variable aléatoire

Y_{ti} = mesure effectuée au temps t chez l'individu i.

- Chaque individu a son propre vecteur de réponses :

$$\mathbf{Y}_i = \begin{pmatrix} Y_{1i} \\ Y_{2i} \\ Y_{3i} \\ \vdots \end{pmatrix} = (Y_{1i}, Y_{2i}, Y_{3i}, \ldots)'$$

Données répétées et longitudinales

- Chaque élément de $\mathbf{Y}_i$ est une variable aléatoire (avec sa moyenne, sa variance, sa loi de probabilité associée).
- Ces variables aléatoires varient ensemble : distribution de probabilité jointe.
- Matrice de covariance :

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_n^2 \end{pmatrix}$$

Difficultés liées aux données répétées et longitudinales

- hétérogénéité non observée (au niveau de l'individu, du groupe) ;
- corrélations résiduelles intra ;
- données absentes ;
- occasions de mesure irrégulièrement espacées ;
- données hiérarchisées.

Gibbons, R., Hedeker, D., & DuToit, S. (2010). Advances in analysis of longitudinal data. *Annual Review of Clinical Psychology*. 6, 79-107.

Modèle à effets fixes et aléatoires pour données balancées

- m individus d'une population, n occasions de mesure.
- $\mathbf{Y}_{ti}$ un vecteur aléatoire dans lequel les variables sont ordonnées à temps croissant :

$$\mathbf{Y}_{ti} = (Y_{1i}, Y_{2i}, \ldots, Y_{ni})'$$

- $\mathbf{Y}_{ti}$ a une distribution normale multivariée de moyenne μ et de variance Σ.
- Les m vecteurs aléatoires $\mathbf{Y}_{i}$ sont indépendants.
- Les n composantes $\mathbf{Y}_{t}$ de $\mathbf{Y}_{ti}$ tendent à être corrélées (Σ n'est sans doute pas diagonale).

Introduction
GLM
MLR
LGC

Modèle à effets fixes et aléatoires pour données balancées

- m vecteurs aléatoires indépendants $\mathbf{Y}_i$:

$$E(\mathbf{Y}_i) = \boldsymbol{\mu} \text{ et } \operatorname{var}(\mathbf{Y}_i) = \boldsymbol{\Sigma}.$$

$$\mathbf{Y}_i = \boldsymbol{\mu} + \boldsymbol{\varepsilon}_i, \; E(\boldsymbol{\varepsilon}_i) = 0, \; \operatorname{var}(\boldsymbol{\varepsilon}_i) = \boldsymbol{\Sigma}.$$

- Les $\boldsymbol{\varepsilon}_i$ sont des vecteurs aléatoires d'écarts à la moyenne : $\boldsymbol{\varepsilon}_i = (\varepsilon_{1i}, \varepsilon_{2i}, \ldots, \varepsilon_{ni})'$.
- Les écarts ε_{ti} sont corrélés, les $\boldsymbol{\varepsilon}_i$ sont indépendants.
- Si les réponses sont continues,

$$\mathbf{Y}_i \sim \mathcal{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \text{ où les } \mathbf{Y}_i \text{ sont indépendants.}$$

Modèle à effets fixes et aléatoires pour données balancées

$$\mathbf{Y}_i = \boldsymbol{\mu} + \boldsymbol{\varepsilon}_i$$

- Approche « population-averaged » : caractérisation des vecteurs de moyennes.

Modèle à effets fixes et aléatoires pour données balancées

- Le vecteur aléatoire $\bar{\mathbf{Y}} = (\bar{Y}_{.1}, \bar{Y}_{.2}, ..., \bar{Y}_{.n})'$ est un estimateur non biaisé de la moyenne $\boldsymbol{\mu}$: $E(\bar{\mathbf{Y}}) = \boldsymbol{\mu}$
- $s_j^2 = \dfrac{\sum_{j=1}^{m} (Y_{ij} - \bar{Y}_{.j})^2}{m-1}$ est un estimateur non biaisé de la variance σ_j^2.
- $s_{jk} = \dfrac{\sum_{i=1}^{m} (Y_{ij} - \bar{Y}_{.j})(Y_{ij} - \bar{Y}_{.k})}{m-1}$ est un estimateur non biaisé de la covariance σ_{jk}.

Modèle à effets fixes et aléatoires pour données balancées

Introduction
GLM
MLR
LGC

- L'estimateur de la matrice de covariance $\boldsymbol{\Sigma}$ est la matrice de covariance :

$$\hat{\boldsymbol{\Sigma}} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_n^2 \end{bmatrix}$$

avec $t = 1,2,\ldots n$ occasions de mesure.

Modèle à effets fixes et aléatoires pour données balancées

Introduction
GLM
MLR
LGC

- L'estimateur de la matrice de corrélation Γ associée à la matrice de covariance est la matrice de corrélation :

$$\hat{\mathbf{\Gamma}} = \begin{bmatrix} 1 & \rho_{12} & \dots & \rho_{1n} \\ \rho_{21} & 1 & \dots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n1} & \rho_{n2} & \dots & 1 \end{bmatrix}$$

- Ce modèle de corrélation est parfois plus facile à discuter en premier.

Introduction
GLM
MLR
LGC

Analyse inter → analyse inter+intra

$$Y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + e_i,$$

$$\downarrow$$

$$Y_{ti} = \beta_0 + \beta_1 x_{ti} + \beta_2 z_i + u_i + e_{ti},$$

Covariable variant dans le temps

Covariable ne variant pas dans le temps

t occasions de mesure

GLM : UANOVA à mesures répétées

Plan simple intra : Y_{ti} = observation de l'individu i au temps t :

$$Y_{ti} = \mu + \gamma_t + u_i + e_{ti}$$

μ : moyenne totale

γ_t : écart associé au temps,

u_i : effet aléatoire (variations entre individus: par hypothèse les mêmes à tous les temps),

e_{ti}: écart aléatoire (variations intra + erreur de mesure).

NB: les valeurs du temps ne figurent pas explicitement dans l'équation du modèle.

GLM : UANOVA à mesures répétées

Hypothèses

- $u_i \sim \mathcal{N}(0, \sigma_u^2)$ et indépendantes ;
- $e_{ti} \sim \mathcal{N}(0, \sigma_e^2)$ et indépendantes ; composante de variance σ_e^2 constante au cours du temps.

Avec ces hypothèses, $\mathbf{Y}_i$ a une distribution normale multivariée de moyenne μ .

GLM : UANOVA à mesures répétées

- Matrice de covariance Σ :

$$\operatorname{var}(\mathbf{Y}_i) = \begin{bmatrix} \sigma_u^2 + \sigma_e^2 & \sigma_u^2 & \cdots & \sigma_u^2 \\ \sigma_u^2 & \sigma_u^2 + \sigma_e^2 & \cdots & \sigma_u^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_u^2 & \sigma_u^2 & \cdots & \sigma_u^2 + \sigma_e^2 \end{bmatrix}$$

- Structure de covariance dite à symétrie composée (*CS*) ou échangeable.

- Corrélation intra-classe $\dfrac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2}$.

GLM : UANOVA à mesures répétées

Plan mixte inter/intra : Y_{thl} = observation au temps t de l'individu h du groupe l :

$$Y_{thl} = \mu + \tau_l + \gamma_t + (\tau\gamma)_{tl} + u_{hl} + e_{thl}$$

μ : moyenne totale

τ_l : écart à la moyenne associé au groupe,

γ_t : écart à la moyenne associé au temps,

$(\tau\gamma)_{tl}$: écart à la moyenne associé à l'interaction groupe*temps

u_{hl} : effet aléatoire (variations entre individus dans les groupes),

e_{thl}: écart aléatoire (variations intra + erreur de mesure).

Introduction
GLM
MLR
LGC

GLM : modèle général de régression

$$Y_{thl} = \boxed{\mu + \tau_l + \gamma_t + (\tau\gamma)_{tl}} + \boxed{u_{hl} + e_{thl}}$$

Facteur systématique — Variations aléatoires

$$\mu_{tl} \qquad \varepsilon_{thl}$$

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\varepsilon}_i$$

β : vecteur colonne contenant les μ, τ_l, γ_t et $(\tau\gamma)_{tl}$ « empilés » (format univarié);

$\mathbf{X}_i$ matrice plan de 0 et de 1.

Plan simple, un groupe

$$\mathbf{Y}_i = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}_i$$

$$\begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_n \end{bmatrix} \times \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} \beta_0 + \beta_1 t_1 \\ \beta_0 + \beta_1 t_2 \\ \vdots \\ \beta_0 + \beta_1 t_n \end{bmatrix}$$

$$\operatorname{var}(\boldsymbol{\varepsilon}_i) = \boldsymbol{\Sigma}.$$

$$\mathbf{Y}_i \sim \mathcal{N}_n(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma})$$

Plan mixte, deux groupes

$$\mathbf{Y}_i = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}_i$$

$$\begin{bmatrix} (1-\delta_i) & (1-\delta_i)t_1 & \delta_i & \delta_i t_1 \\ (1-\delta_i) & (1-\delta_i)t_2 & \delta_i & \delta_i t_2 \\ \vdots & \vdots & \vdots & \vdots \\ (1-\delta_i) & (1-\delta_i)t_n & \delta_i & \delta_i t_n \end{bmatrix} \times \begin{bmatrix} \beta_{0G} \\ \beta_{1G} \\ \beta_{0F} \\ \beta_{1F} \end{bmatrix} = \begin{bmatrix} (1-\delta_i)\beta_{0G} + (1-\delta_i)\beta_{1G}t_1 + \delta_i\beta_{0F} + \delta_i\beta_{1F}t_1 \\ (1-\delta_i)\beta_{0G} + (1-\delta_i)\beta_{1G}t_2 + \delta_i\beta_{0F} + \delta_i\beta_{1F}t_2 \\ \vdots \\ (1-\delta_i)\beta_{0G} + (1-\delta_i)\beta_{1G}t_n + \delta_i\beta_{0F} + \delta_i\beta_{1F}t_n \end{bmatrix}$$

$$\begin{aligned} \text{avec } \delta_i &= 0 \text{ si } i = 0 \\ \delta_i &= 1 \text{ si } i = 1. \end{aligned}$$

$$\mathrm{var}(\boldsymbol{\varepsilon}_i) = \boldsymbol{\Sigma}.$$

$$\mathbf{Y}_i \sim \mathcal{N}_n(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma})$$

Plan mixte, deux groupes, effet quadratique du temps, une covariable

$$\mathbf{Y}_i = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}_i$$

$$\begin{bmatrix} (1-\delta_i) & (1-\delta_i)t_1 & (1-\delta_i)t_1^2 & \delta_i & \delta_i t_1 & \delta_i t_1^2 & a_i \\ (1-\delta_i) & (1-\delta_i)t_2 & (1-\delta_i)t_2^2 & \delta_i & \delta_i t_2 & \delta_i t_2^2 & a_i \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ (1-\delta_i) & (1-\delta_i)t_n & (1-\delta_i)t_2^2 & \delta_i & \delta_i t_n & \delta_i t_n^2 & a_i \end{bmatrix} \times \begin{bmatrix} \beta_{0G} \\ \beta_{1G} \\ \beta_{2G} \\ \beta_{0F} \\ \beta_{1F} \\ \beta_{2F} \\ \beta_3 \end{bmatrix}$$

avec $\delta_i = 0$ si $i = 0$

$\delta_i = 1$ si $i = 1$.

$$\mathrm{var}(\boldsymbol{\varepsilon}_i) = \boldsymbol{\Sigma}.$$

$$\mathbf{Y}_i \sim \mathcal{N}_n(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma})$$

GLM : UANOVA à mesures répétées

- Hypothèses de la matrice de covariance
  - Hypothèse d'homogénéité des variances (homoscédasticité)
  - Hypothèse d'égalité de covariances (CS)
  - Si CS ne tient pas, hypothèse plus souple dite de sphéricité (matrice de type H: variance constante des différences de scores).
- Risque d'erreurs de type I (H_0 trop souvent rejetée) si ces hypothèses ne sont pas respectées.

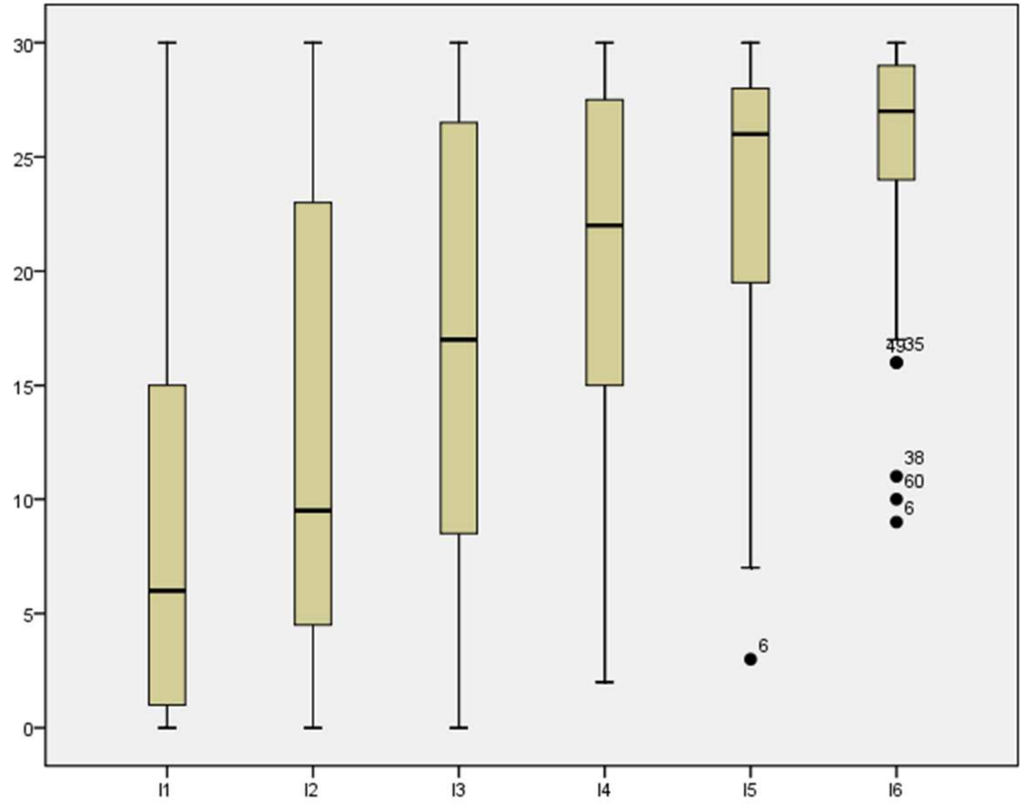
GLM : MANOVA à mesures répétées

- Hypothèse d'une matrice de covariance **non structurée** des mesures répétées.
- Mais :
  - Nécessite des données complètes.
  - Occasions de mesure (*time-points*) fixées et traitées comme une variable de classification (*factor*).
  - Modélisation possible seulement pour des covariables ne variant pas dans le temps (*TIC*).

Introduction
GLM
MLR
LGC

Illustration : plan simple intra

- Données lecture_all (6 mesures de lecture de mots, données balancées, N=60)



GLM : plan simple intra

Code SPSS

```
GLM l1 l2 l3 l4 l5 l6 (fichier lecture_wide)
  /WSFACTOR=session 6 Polynomial
  /METHOD=SSTYPE(3)
  /EMMEANS=TABLES(session)
  /PRINT=PARAMETER RSSCP
  /CRITERIA=ALPHA(.05)
  /WSDESIGN=session
```

Hypothèse de sphéricité non respectée

Effets linéaire et quadratique du temps

GLM : plan simple intra

Parameter Estimates

Dependent Variable	Parameter	B	Std. Error	t	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
I1	Intercept	9,200	1,153	7,978	,000	6,892	11,508
I2	Intercept	12,767	1,254	10,177	,000	10,257	15,277
I3	Intercept	16,533	1,209	13,677	,000	14,114	18,952
I4	Intercept	20,567	1,038	19,812	,000	18,489	22,644
I5	Intercept	23,400	,851	27,504	,000	21,698	25,102
I6	Intercept	25,583	,654	39,112	,000	24,274	26,892

- Matrice de covariance résiduelle :

		I1	I2	I3	I4	I5	I6
Covariance	I1	79,790	77,217	69,942	52,478	39,834	25,153
	I2	77,217	94,419	82,093	63,084	44,807	29,918
	I3	69,942	82,093	87,677	66,659	51,241	33,040
	I4	52,478	63,084	66,659	64,656	47,634	32,596
	I5	39,834	44,807	51,241	47,634	43,431	28,831
	I6	25,153	29,918	33,040	32,596	28,831	25,671
Correlation	I1	1,000	,890	,836	,731	,677	,556
	I2	,890	1,000	,902	,807	,700	,608
	I3	,836	,902	1,000	,885	,830	,696
	I4	,731	,807	,885	1,000	,899	,800
	I5	,677	,700	,830	,899	1,000	,863
	I6	,556	,608	,696	,800	,863	1,000

MIXED : plan simple intra

La procédure MIXED (SPSS) permet d'estimer les effets fixes et aléatoires avec différentes hypothèses sur la matrice de covariance.

MIXED lect BY session (fichier lecture_long)

/CRITERIA=CIN(95) MXITER(1000) MXSTEP(5) SCORING(1)
SINGULAR(0.000000000001) HCONVERGE(0, ABSOLUTE)
LCONVERGE(0, ABSOLUTE) PCONVERGE(0.000001, ABSOLUTE)

/**FIXED=session** | SSTYPE(3)

/METHOD=ML

/PRINT=R SOLUTION

/**REPEATED=session** | SUBJECT(id) **COVTYPE(UN)**.

BIC = 2167

Introduction
GLM
MLR
LGC

MIXED : plan simple intra

Paramètres estimés

Estimates of Fixed Effects[b]

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Intercept	25,583333	,648627	60,000	39,442	,000	24,285885	26,880781
[session=1]	-16,383333	,950757	60,000	-17,232	,000	-18,285131	-14,481535
[session=2]	-12,816667	,993728	60	-12,898	,000	-14,804419	-10,828915
[session=3]	-9,050000	,880159	60,000	-10,282	,000	-10,810581	-7,289419
[session=4]	-5,016667	,641825	60,000	-7,816	,000	-6,300509	-3,732825
[session=5]	-2,183333	,433007	60,000	-5,042	,000	-3,049477	-1,317190
[session=6]	0[a]	0	.	.	.	.	.

a. This parameter is set to zero because it is redundant.

b. Dependent Variable: lect.

Matrice de covariance résiduelle

Residual Covariance (R) Matrix[a]

	[session = 1]	[session = 2]	[session = 3]	[session = 4]	[session = 5]	[session = 6]
[session = 1]	78,460000	75,930000	68,776667	51,603333	39,170000	24,733333
[session = 2]	75,930000	92,845556	80,724444	62,032222	44,060000	29,419444
[session = 3]	68,776667	80,724444	86,215556	65,547778	50,386667	32,488889
[session = 4]	51,603333	62,032222	65,547778	63,578889	46,840000	32,052778
[session = 5]	39,170000	44,060000	50,386667	46,840000	42,706667	28,350000
[session = 6]	24,733333	29,419444	32,488889	32,052778	28,350000	25,243056

Unstructured

a. Dependent Variable: lect.

Illustration : plan simple intra

MIXED

D'autres matrices de covariance peuvent être spécifiées :

Σ	#par	BIC
Diagonale	6	2561
Symétrie Composée	2	2245
AR(1)	2	2119
Identité	1	2565
Toeplitz	6	2132
Non structurée	21	2167

Illustration : plan mixte inter/intra

- Données orthodont (croissance faciale à 8, 10, 12 et 14 ans : 16 garçons, 11 filles)

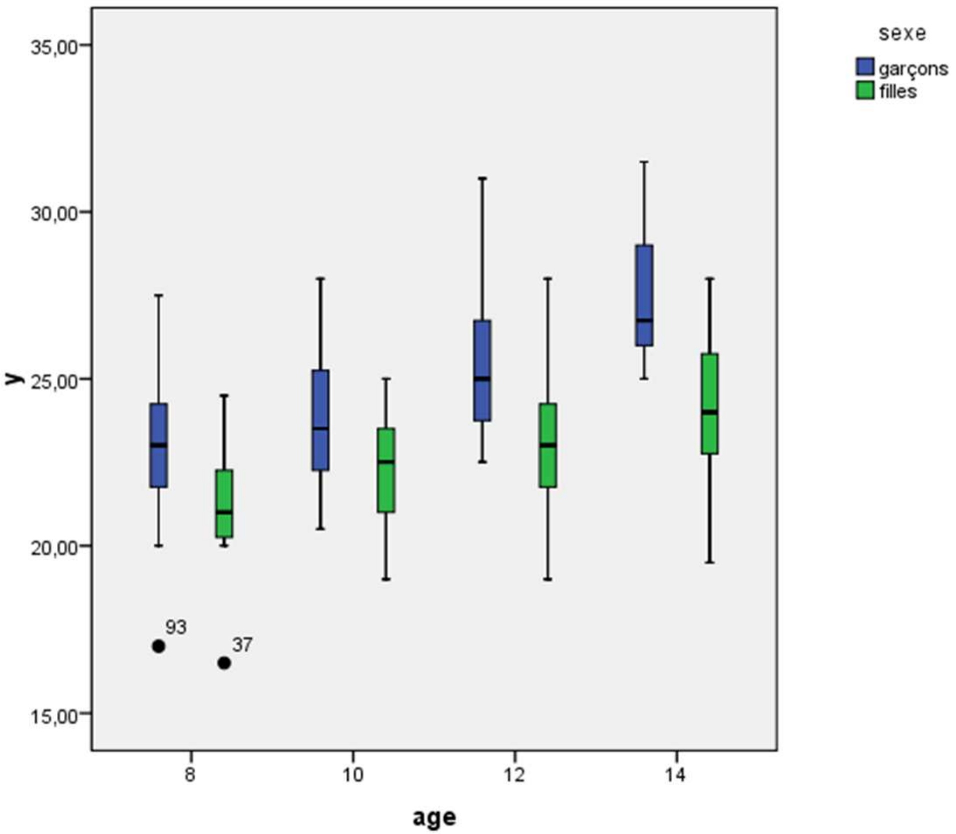
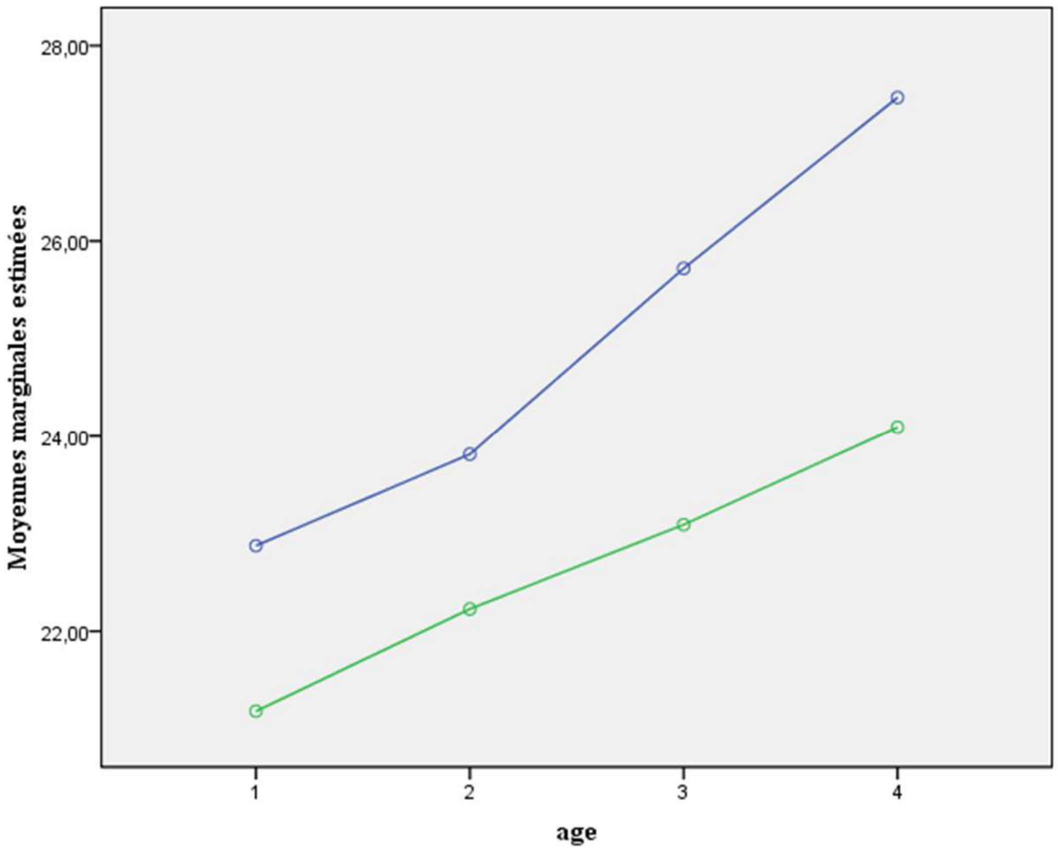


Illustration : plan mixte inter/intra

- Données orthodont : moyennes marginales estimées



Introduction
GLM
MLR
LGC

Illustration : plan mixte inter/intra

- Données orthodont_wide : matrices de covariance et de corrélation pour chaque groupe

	age1	age2	age3	age4
age1	6,017	2,292	3,629	1,613
age2	2,292	4,563	2,194	2,810
age3	3,629	2,194	7,032	3,241
age4	1,613	2,810	3,241	4,349

	age1	age2	age3	age4
age1	1	,437	,558*	,315
age2	,437	1	,387	,631**
age3	,558*	,387	1	,586*
age4	,315	,631**	,586*	1

garçons

	age1	age2	age3	age4
age1	4,514	3,355	4,332	4,357
age2	3,355	3,618	4,027	4,077
age3	4,332	4,027	5,591	5,466
age4	4,357	4,077	5,466	5,941

	age1	age2	age3	age4
age1	1	,830**	,862**	,841**
age2	,830**	1	,895**	,879**
age3	,862**	,895**	1	,948**
age4	,841**	,879**	,948**	1

filles

Illustration : plan mixte inter/intra

```
MIXED BY sexe age
  /CRITERIA=CIN(95) MXITER(100) MXSTEP(5) SCORING(1)
    SINGULAR(0.000000000001) HCONVERGE(0, ABSOLUTE)
    LCONVERGE(0, ABSOLUTE) PCONVERGE(0.000001, ABSOLUTE)
  /FIXED=sexe age sexe*age | SSTYPE(3)
  /METHOD=ML
  /PRINT=R SOLUTION
  /REPEATED=age | SUBJECT(id) COVTYPE(UN).
```

BIC = 501

Introduction
GLM
MLR
LGC

Illustration : plan mixte inter/intra

MIXED

D’autres matrices de covariance peuvent être spécifiées :

Σ	#par	BIC	sexe*age
Diagonale	4	532	p=0,386
Compound Symmetry	2	473	p=0,061
AR(1)	2	485	p=0,310
Identité	1	520	p=0,416
Toeplitz	6	478	p=0,069
Non structurée	21	501	p=0,041

Cf . Séminaire 27 janvier 2011 Tours\analyses SPSS\SPSS_Syntax_Part_3

Inconvénients et limites de l'ANOVA à mesures répétées

- Les individus dévient de leur pattern de réponse moyen par un même écart, constant au cours du temps.
- Données balancées (même nombre de mesures pour tous les sujets, mêmes temps d'observation) ;
- Temps = facteur catégoriel (intervalles de temps égaux).
- Même matrice de covariance pour tous les vecteurs de données.
- Hypothèses sur la matrice de covariance : très restrictive (CS) pour l'ANOVA; très libérale (UN) pour la MANOVA).
- Covariables temporellement invariantes (pour MANOVA).

Based on these limitations, the repeated measures ANOVA and related approaches should no longer be used for analysis of longitudinal data.

(Gibbons, Hedeker & DuToit, 2010)

2. Modèles de courbe de croissance

Deux grandes approches

- Régression multi-niveaux (*MLR*) : R_nlme, R_lme4, SAS PROC MIXED, SPSS MIXED, Mplus,…
  - Les mesures répétées (niveau 1) sont emboîtées dans chaque individu (niveau 2).
- Approche par les modèles structuraux (*SEM*) : R_sem, LISREL, AMOS, Mplus,...
  - Les mesures répétées sont les indicateurs multiples d'un ou de plusieurs facteurs de croissance caractérisant les trajectoires latentes.

Introduction
GLM
MLR
LGC

2.a. Régression multi-niveaux (MLR)

Bressoux, P. (2008). *Modélisation statistique appliquée aux sciences sociales* (Les modèles de croissance: pp. 359-405). Bruxelles: De Boeck Université.

- Modèles statistiques pour données répétées ou longitudinales permettant d'estimer la variabilité interindividuelle dans le changement intra-individuel :
  - Modèles linéaires mixtes, modèles à effets mixtes, modèles des composant(e)s de la variance, modèles à coefficients aléatoires, modèles linéaires hiérarchiques.
- Temps : niveau 1; individu : niveau 2;
- Prédicteurs pouvant être introduits à tous les niveaux.
- Les DI dans les paramètres qui décrivent la courbe de croissance de la VD en f° du temps sont modélisés comme des effets aléatoires.

Modèle statistique de base

Approche « subject-specific »: caractérisation de la nature des écarts aléatoires.

$$Y_{ti} = \mu_t + \varepsilon_{ti} = \mu_t + u_{ti} + e_{ti1} + e_{ti2}$$

variations interindividuelles

variations intra-individuelles

erreur de mesure

Description des DI dans le changement (effets aléatoires): les variances et les covariances peuvent différer au cours du temps.

Les effets aléatoires apparaissent comme un moyen de structurer la matrice de variance-covariance des observations.

Introduction
GLM
MLR
LGC

Modèle vide

Modèle pour les moyennes – Modèle pour les variances

$$Y_{ti} = \beta_{0i} + e_{ti}$$
$$\beta_{0i} = \gamma_{00} + u_{0i}$$

1 effet fixe: intercept γ_{00}

1 effet aléatoire (intercept): u_{0i} (between)

1 erreur résiduelle : e_{ti} (within)

Modèle à effets aléatoires du temps

$$Y_{ti} = \beta_{0i} + \beta_{1i} \times Temps_{ti} + e_{ti}$$

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

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

soit :

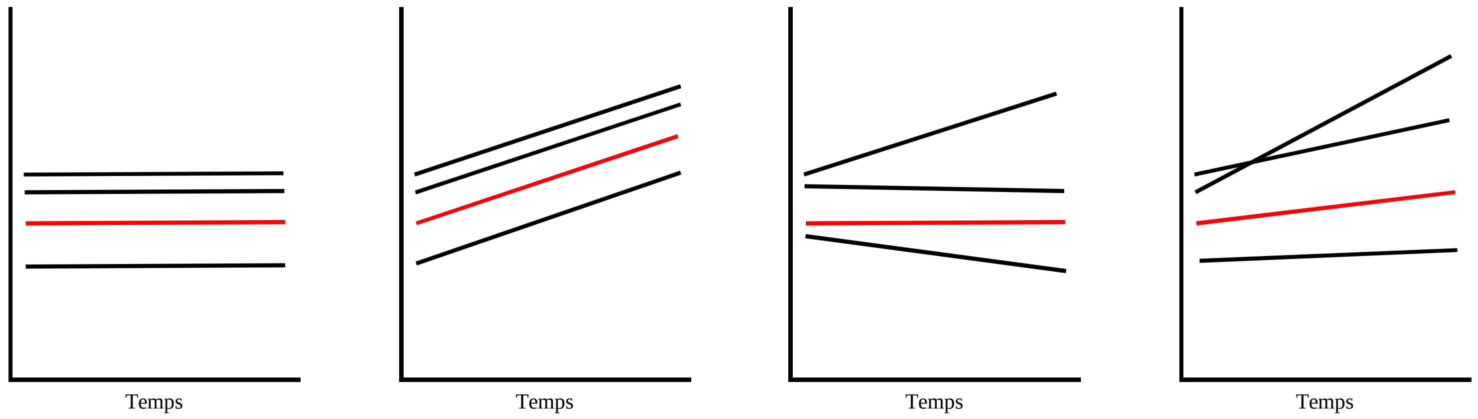
$$Y_{ti} = (\gamma_{00} + u_{0i}) + (\gamma_{10} + u_{1i}) \times Temps_{ti} + e_{ti}$$

2 effets fixes: intercept γ_{00} et pente γ_{10}

2 (+1) effets aléatoires : variances de u_{0i} et u_{1i}, $\mathrm{cov}(u_{0i}, u_{1i})$

1 erreur résiduelle : e_{ti}

Modèle à effets aléatoires du temps



- Pas d’effet fixe
- Pas d’effet aléatoire

- Effet principal individu

- Effet fixe
- Pas d’effet aléatoire

- Effet principal individu et temps

- Pas d’effet fixe
- Effet aléatoire

- Effet individu*temps
- Pas d’effet principal du temps

- Effet fixe
- Effet aléatoire

- Effet individu*temps
- Effet principal du temps

Introduction
GLM
MLR
LGC

En résumé

Variance résiduelle totale = 3 composantes

- Variance inter autour de l'intercept
- Variance inter autour de la pente
- Variance résiduelle intra

Variance résiduelle ANOVA

Réduction de la variance résiduelle totale par introduction dans le modèle de prédicteurs de niveau :

- individuel ou inter (between) : TIC (*time invariant covariates*)
- temporel ou intra (within): TVC (*time-varying covariates*)

Estimation des paramètres

Fonction de vraisemblance

Probabilité d'observer un échantillon de données sachant les paramètres du modèle (effets fixes, effets aléatoires, variance résiduelle). La vraisemblance (produit de probabilités) a une expression dérivable par rapport au vecteur de paramètres θ.

Estimateur du maximum de vraisemblance à information complète (*full ML*)

Maximisation de la vraisemblance des données d'échantillon (ajustement des effets fixes et aléatoires du modèle).

Estimateur du maximum de vraisemblance restreint (*REML*)

Maximisation de la vraisemblance des résidus d'échantillon (après estimation par MCO ou généralisés des effets fixes) afin de corriger la sous-estimation des effets aléatoires par ML.

Comparaison de modèles

Déviance $= -2\left[LL_{\text{modèle testé}} - LL_{\text{modèle saturé}}\right]$

Déviance = indice de « mauvais » ajustement

Attention : ML : ajustement du modèle pour les moyennes et du modèle pour les variances *vs* REML: ajustement du modèle pour les variances.

Critères d'information

AIC = deviance + 2 (#par)

BIC = deviance + ln N (#par)

Comparaison de modèles

Modèles emboîtés

Effets fixes : *p* ou test de différence de déviance (ML)

Effets aléatoires : test de différence de déviance (REML)

Effets fixes et aléatoires : test de différence de déviance (ML)

Modèles non emboîtés

Effets fixes : ML, BIC

Effets aléatoires: REML, BIC

Effets fixes et aléatoires: ML, BIC

Comparaison de modèles

Test de différence de déviance

chi-deux = $\text{Deviance}_{\text{baseline}} - \text{Deviance}_{\text{alternatif}}$

ddl = $\#\text{par}_{\text{baseline}} - \#\text{par}_{\text{alternatif}}$

Démarche

- Commencer avec REML : effets fixes en fonction de la valeur de *p*;

- Tests de différence de déviance : introduction d'effets aléatoires.

Structure de covariance : matrices employées (écriture sous SPSS, SAS)

Matrice R de covariance intra-individuelle

Niveau 1 : e_{ti}

Matrice G de covariance des effets aléatoires (inter)

Niveau 2 : u_i

$$\mathbf{V} = \begin{bmatrix} \sigma_{11}^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22}^2 & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn}^2 \end{bmatrix} = \mathbf{G} \,\&\, \mathbf{R}$$

Structure de covariance : matrices employées (écriture sous SPSS, SAS)

Introduction
GLM
MLR
LGC

Exemple : $$Y_{ti} = (\gamma_{00} + u_{0i}) + e_{ti}$$

1) Modèle à intercept aléatoire :

- G : variance de l'intercept (RANDOM)
- R: matrice identité (REPEATED)

2) Modèle à symétrie composée

- G: rien
- R: matrice à symétrie composée (REPEATED)

Structure de covariance : matrices employées (écriture sous SPSS, SAS)

Exemple : $Y_{ti} = (\gamma_{00} + u_{0i}) + e_{ti}$

1) Modèle à intercept aléatoire :

```
MIXED lect WITH session
/PRINT=G  R SOLUTION
  /RANDOM=INTERCEPT | SUBJECT(id) COVTYPE(UN)
  /REPEATED=session | SUBJECT(id) COVTYPE(ID). !COVTYPE(DIAG)
```

2) Modèle à symétrie composée :

```
MIXED lect WITH session
/PRINT=G  R SOLUTION
  /REPEATED=session | SUBJECT(id) COVTYPE(CS).
```

Introduction
GLM
MLR
LGC

Structure de covariance : matrices employées (écriture sous SPSS, SAS)

Exemple :

Estimates of Fixed Effects[a]

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Intercept	18,008333	,934100	59	19,279	,000	16,139204	19,877463

a. Dependent Variable: lect.

1)

Random Effect Covariance Structure (G)[a]

	Intercept | id
Intercept | id	43,005807

Variance Components

a. Dependent Variable: lect.

&

Residual Covariance (R) Matrix[a]

	[session = 1]	[session = 2]	[session = 3]	[session = 4]	[session = 5]	[session = 6]
[session = 1]	56,080556	0	0	0	0	0
[session = 2]	0	56,080556	0	0	0	0
[session = 3]	0	0	56,080556	0	0	0
[session = 4]	0	0	0	56,080556	0	0
[session = 5]	0	0	0	0	56,080556	0
[session = 6]	0	0	0	0	0	56,080556

Identity

a. Dependent Variable: lect.

2)

Residual Covariance (R) Matrix[a]

	[session = 1]	[session = 2]	[session = 3]	[session = 4]	[session = 5]	[session = 6]
[session = 1]	99,086362	43,005807	43,005807	43,005807	43,005807	43,005807
[session = 2]	43,005807	99,086362	43,005807	43,005807	43,005807	43,005807
[session = 3]	43,005807	43,005807	99,086362	43,005807	43,005807	43,005807
[session = 4]	43,005807	43,005807	43,005807	99,086362	43,005807	43,005807
[session = 5]	43,005807	43,005807	43,005807	43,005807	99,086362	43,005807
[session = 6]	43,005807	43,005807	43,005807	43,005807	43,005807	99,086362

Compound Symmetry

a. Dependent Variable: lect.

Introduction
GLM
MLR
LGC

Modèle linéaire général mixte

$$\mathbf{Y}_{ti} = \mathbf{X}_{ti}\boldsymbol{\beta} + \mathbf{Z}_{ti}\mathbf{u}_i + \mathbf{e}_{ti}$$

$$\mathbf{u}_i \sim \mathcal{N}(0, \mathbf{G})$$

$$\mathbf{e}_i \sim \mathcal{N}(0, \sigma_e^2 \mathbf{I}_{ti})$$

$u_1, u_2, \ldots, u_n, e_1, e_2, \ldots e_n$ indépendants.

Effets fixes: β

Effets aléatoires: $\mathbf{u}_i$

Composant(e)s de la variance: éléments de G et R

Introduction
GLM
MLR
LGC

Deux familles de modèles pour les variances

Modèles mixtes : mise en compétition de diverses matrices de covariance

Pas de changement intra- dans le temps mais les variances et covariances diffèrent au cours du temps.

Structuration de R (REPEATED)

Modèles multi-niveaux

Changement intra- dans le temps.

Intercept aléatoire? Pente aléatoire? Intercept et pente aléatoires?

Structuration de G (RANDOM) avec $R=\sigma^2 I$

Stratégie de modélisation

- Modèle vide (ICC) : intercept aléatoire dans G, R=I (ou rien dans G, R=symétrie composée).
- Choix d'une métrique pour le temps ; centrer pour faciliter l'interprétation des coefficients des effets fixes
- **Changement systématique de la VD**: RANDOM

  Effet linéaire (quadratique, etc.) fixe, intercept et pente linéaire (quadratique, etc.) aléatoires: effets aléatoires dans G (non structurée), R=I.

  → Avantage: séparation entre G et R
- **Pas de changement systématique de la VD**: REPEATED – rien dans G, structure spécifique de toute la variance dans R

  → Inconvénient: pas de séparation entre G et R.

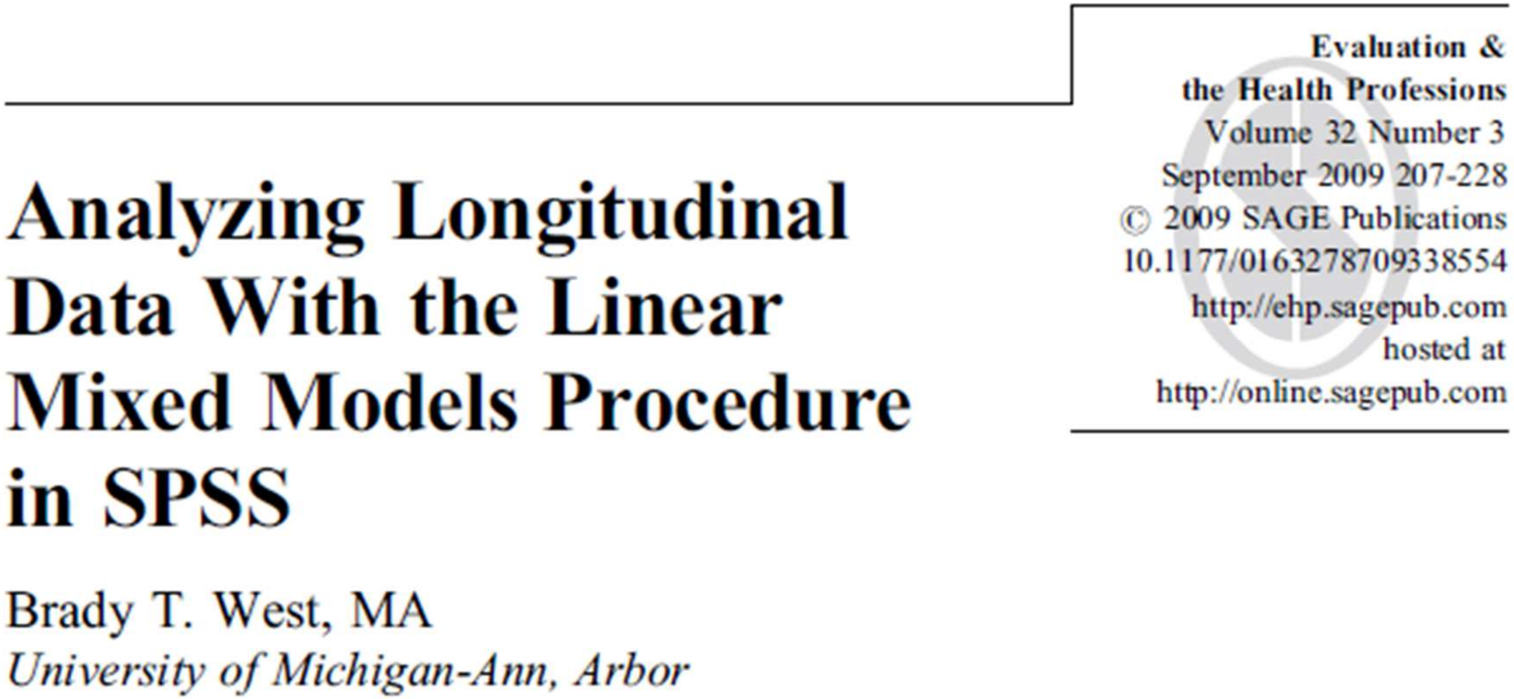
Stratégie de modélisation

- Description préalable des données
- Modèle vide
- Choisir t=0 (*centering time-point**)
- Estimation (ML ou REML) des moyennes (surtout si données manquantes)
- Evaluation des effets fixes et aléatoires du temps (focus sur G)
- Evaluation d'autres matrices de covariance (focus sur R), comparaison d'autres modèles avec les mêmes effets fixes.

Biesanz, J., *et al.* (2004). The role of coding time in estimating and interpreting growth curve models. *Psychological Methods, 9*, 30-52.

Introduction
GLM
MLR
LGC

Guide pratique



Evaluation & the Health Professions
Volume 32 Number 3
September 2009 207-228

10.1177/0163278709338554
http://ehp.sagepub.com
hosted at
http://online.sagepub.com

Analyzing Longitudinal Data With the Linear Mixed Models Procedure in SPSS

Brady T. West, MA
University of Michigan-Ann, Arbor

Introduction
GLM
MLR
LGC

Illustration : MLR avec SPSS

- Fichier Tours_long_2011.sav

```
*ICC*
MIXED lect
/FIXED=intercept
  /METHOD=ML
 /RANDOM=INTERCEPT | SUBJECT(id1) COVTYPE(UN)
  /PRINT = solution testcov.
```

Estimates of Fixed Effects[a]

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	19,770000	,914474	60	21,619	,000	17,940780	21,599220

a. Dependent Variable: lect.

Estimates of Covariance Parameters[a]

Parameter		Estimate	Std. Error	Wald Z	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Residual		41,710000	3,807585	10,954	,000	34,876788	49,882005
Intercept [subject = id1]	Variance	41,833767	9,192397	4,551	,000	27,194877	64,352710

a. Dependent Variable: lect.

Illustration : MLR avec SPSS

- Fichier Tours_long_2011.sav

```
*croissance linéaire*
MIXED lect WITH session
/FIXED=intercept session | SSTYPE(3)
  /METHOD=ML
  /PRINT=G  R SOLUTION
  /RANDOM=INTERCEPT session | SUBJECT(id1) COVTYPE(UN).
```

Introduction
GLM
MLR
LGC

Estimates of Fixed Effects[a]

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	10,020000	1,510393	60,000	6,634	,000	6,998764	13,041236
session	3,250000	,256661	60,000	12,663	,000	2,736601	3,763399

a. Dependent Variable: lect.

Random Effect Covariance Structure (G)[a]

	Intercept | id1	session | id1
Intercept | id1	128,924267	-18,210000
session | id1	-18,210000	3,229500

Unstructured

a. Dependent Variable: lect.

Residual Covariance (R) Matrix[a]

	Residual
Residual	7,230000

a. Dependent Variable: lect.

Illustration : MLR avec SPSS

- Fichier Tours_long_2011.sav

```
*croissance quadratique*
MIXED lect WITH session session2
/FIXED=intercept session session2| SSTYPE(3)
 /METHOD=ML
 /PRINT=G  R SOLUTION
 /RANDOM=INTERCEPT session session2 | SUBJECT(id1) COVTYPE(UN).
```

Estimates of Fixed Effects[a]

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	7,836667	1,578392	60,000	4,965	,000	4,679412	10,993921
session	5,121429	,739174	60,000	6,929	,000	3,642860	6,599997
session2	-,311905	,111460	60,000	-2,798	,007	-,534858	-,088951

a. Dependent Variable: lect.

Random Effect Covariance Structure (G)[a]

	Intercept | id1	session | id1	session2 | id1
Intercept | id1	126,726846	-26,275071	1,035960
session | id1	-26,275071	19,569310	-2,518929
session2 | id1	1,035960	-2,518929	,392103

Unstructured

a. Dependent Variable: lect.

Residual Covariance (R) Matrix[a]

	Residual
Residual	4,946190

a. Dependent Variable: lect.

Illustration : MLR avec SPSS

- Fichier Tours_long_2011.sav

```
*croissance quadratique, TIC: PEPH1 etc.*
MIXED lect WITH session session2 PEPH1 etc.
/FIXED=intercept session session2 PEPH1 etc. | SSTYPE(3)
  /METHOD=ML
  /PRINT=G  R SOLUTION
  /RANDOM=INTERCEPT session session2 | SUBJECT(id1) COVTYPE(UN).
```

Introduction
GLM
MLR
LGC

Estimates of Fixed Effects[a]

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	1,211026	2,475572	92,253	,489	,626	-3,705495	6,127547
session	5,121429	,739174	60,000	6,929	,000	3,642860	6,599997
session2	-,311905	,111460	60,000	-2,798	,007	-,534858	-,088951
PEPH1	,958898	,264797	60,000	3,621	,001	,429225	1,488571
PISyl1	,543161	,354619	60,000	1,532	,131	-,166182	1,252505
int1	,116060	,152716	60,000	,760	,450	-,189417	,421538
PIPH1	,160057	,346513	60,000	,462	,646	-,533072	,853187
PESY1	-,169900	,300029	60,000	-,566	,573	-,770048	,430247

a. Dependent Variable: lect.

Introduction
GLM
MLR
LGC

Illustration : MLR avec SPSS

- Fichier Tours_long_2011.sav

```
*croissance quadratique, TIC: PEPH1*
MIXED lect WITH session session2 PEPH1
/FIXED=intercept session session2 PEPH1 | SSTYPE(3)
  /METHOD=ML
  /PRINT=G  R SOLUTION
  /RANDOM=INTERCEPT session session2 | SUBJECT(id1) COVTYPE(UN).
```

Estimates of Fixed Effects[a]

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	4,766018	1,500149	71,277	3,177	,002	1,775006	7,757029
session	5,121429	,739174	60,000	6,929	,000	3,642860	6,599997
session2	-,311905	,111460	60,000	-2,798	,007	-,534858	-,088951
PEPH1	1,096660	,207147	60,000	5,294	,000	,682306	1,511015

a. Dependent Variable: lect.

Random Effect Covariance Structure (G)[a]

	Intercept | id1	session | id1	session2 | id1
Intercept | id1	92,089659	-23,582248	1,082177
session | id1	-23,582248	19,569310	-2,518929
session2 | id1	1,082177	-2,518929	,392103

Unstructured

a. Dependent Variable: lect.

Residual Covariance (R) Matrix[a]

	Residual
Residual	4,946190

a. Dependent Variable: lect.

Illustration : MLR avec SPSS

- Fichier Tours_long_2011.sav

```
*croissance quadratique, TIC: PEPH1, TVC: Let*
MIXED lect WITH session session2 PEPH1
/FIXED=intercept session session2 PEPH1 | SSTYPE(3)
  /METHOD=ML
  /PRINT=G  R SOLUTION
  /RANDOM=INTERCEPT session session2 Let | SUBJECT(id1) COVTYPE(UN).
```

Problème de convergence.

2.b. Modélisation structurale (LGC)

Modèle de base

- Temps = valeurs contraintes des saturations de la variable latente (VL) qui représente la pente de la courbe de croissance.
- Les saturations de la VL qui représente l'intercept de la courbe de croissance sont fixées à 1.
- Les moyennes des facteurs de croissance décrivent la croissance moyenne.
- Les DI dans les paramètres qui décrivent la courbe de croissance sont modélisés comme les (co)variances des facteurs de croissance.
- SEM : moyenne et structure de covariance des VL; MLR : effets fixes et aléatoires.

Modèle de courbe de croissance latente

$$y_{ti} = \lambda_{0t}\eta_{0i} + \lambda_{1t}\eta_{1i} + \gamma_{2t}x_{ti} + \varepsilon_{ti}$$
$$\eta_{0i} = \nu_0 + \gamma_0 w_i + \zeta_{0i}$$
$$\eta_{1i} = \nu_1 + \gamma_1 w_i + \zeta_{1i}$$

$\lambda_{0t} = 1$, λ_{1t} temps de mesure (par ex., [0, 1, 2,…]),

η_{0i} intercept individuel, η_{1i} pente individuelle,

ν_0 et ν_1, moyennes de l'intercept et de la pente,

ζ_{0i} et ζ_{1i} , écarts aléatoires à la moyenne de l'intercept et de la pente,

γ_{2t} , effet sur y_i des covariables x_{ti} qui varient dans le temps (TVC),

γ_0 et γ_1, effets sur le niveau initial et la pente des covariables w_i qui ne varient pas dans le temps (TIC),

ε_{ti}, écarts spécifiques au temps avec $\varepsilon_{ti} \sim \mathcal{N}(0, \sigma_\varepsilon^2)$,

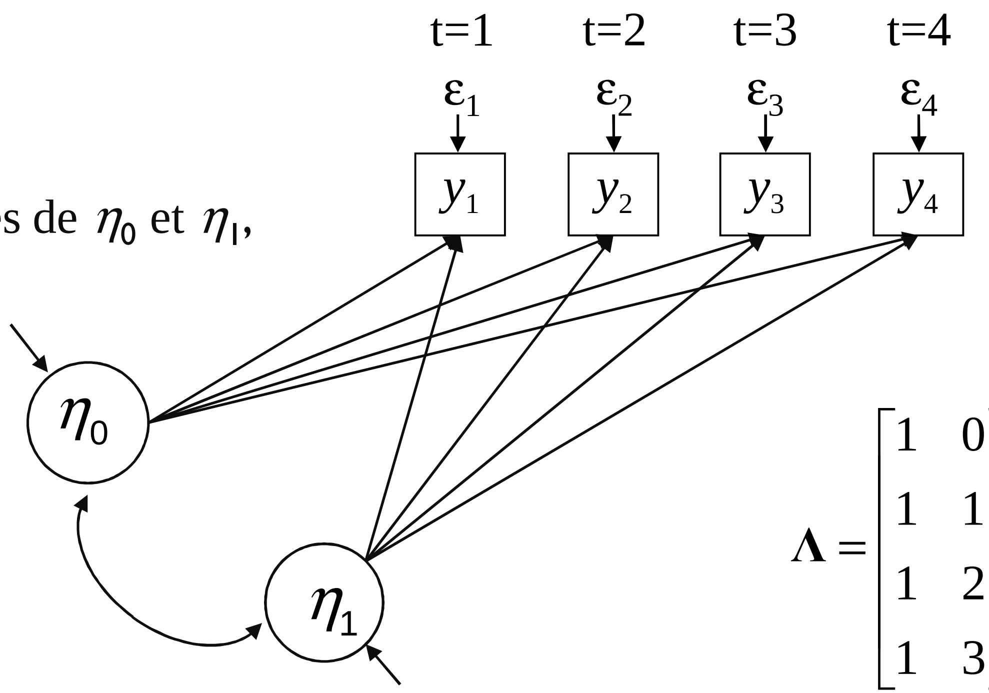
variances et covariances de ζ_{0i} et ζ_{1i} : $\boldsymbol{\Sigma}_\zeta = \begin{bmatrix} \sigma_0^2 & \\ \sigma_{01} & \sigma_1^2 \end{bmatrix}$,

ε_{ti}, η_{0i} et η_{1i} indépendants.

Courbe de croissance à variables latentes : modèle inconditionnel

Paramètres libres (représentation Mplus)

α : moyennes de η_0 et η_1,



$\mathbf{\nu} = \mathbf{0}$

$\mathbf{\Psi}$

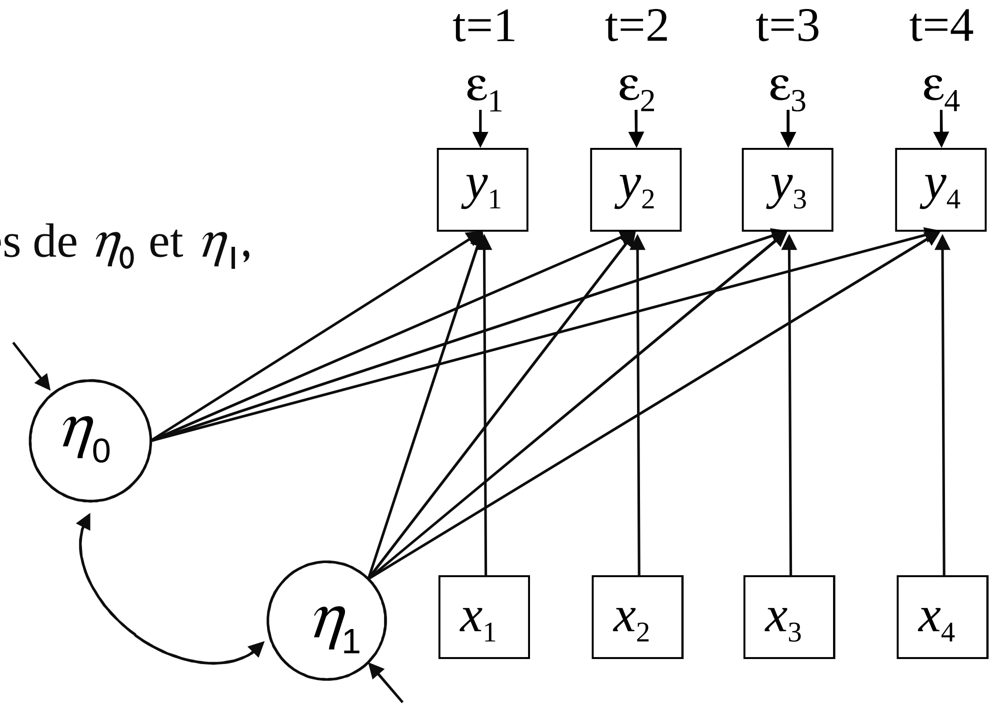
var(η_0)
var(η_1)
cov(η_0,η_1)
var.resid.(y_t)

$$\mathbf{\Lambda} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}$$

$$\mathbf{B} = \mathbf{0}$$

Courbe de croissance à variables latentes : covariables variant dans le temps

Paramètres libres (représentation Mplus)



$\boldsymbol{\nu} = \mathbf{0}$

α : moyennes de η_0 et η_1,

$$\boldsymbol{\Lambda} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}$$

$\boldsymbol{\Psi}$

var(η_0)
var(η_1)
cov(η_0, η_1)
var.resid.(y_t)
cov(x_t, η_0)
cov(x_t, η_1)

B Coefficients de régression de y_t sur x_t

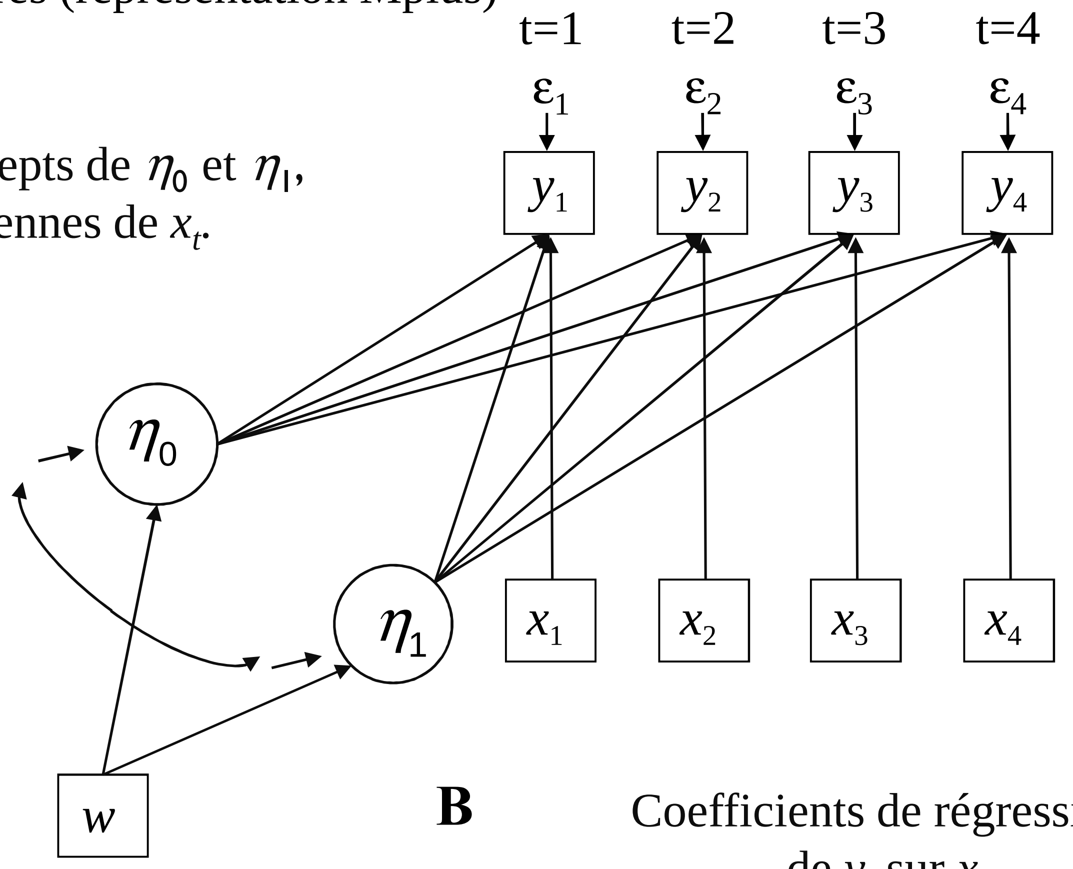
Courbe de croissance à variables latentes : covariables invariantes et variant dans le temps

Paramètres libres (représentation Mplus)

α : intercepts de η_0 et η_1,
moyennes de x_t.

$\nu = \mathbf{0}$

$$\Lambda = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}$$



Ψ
var(η_0)
var(η_1)
cov(η_0, η_1)
var.resid.(y_t)
cov(x_t, η_0)
cov(x_t, η_1)
cov(x_t, w)

Coefficients de régression
de y_t sur x_t
de η_0 et η_1 sur w

Paramètres

- **Facteurs de croissance** :
  - Moyenne et variance de l'intercept
  - Moyenne et variance de la pente
  - Covariance entre intercept et pente
- **Vecteur de réponses $\boldsymbol{y_t}$**
  - Intercepts de y_t fixés à 0 (invariance de mesure)
  - Variances résiduelles: spécifiques au temps et erreur de mesure
  - Covariances résiduelles: entre sources de variation dans le temps
- **Covariables**
  - Coefficients de régression des y_t sur les x_t, des facteurs de croissance sur w.

Mplus : estimation et évaluation de l'ajustement du modèle

- **Plusieurs méthodes d'estimation**
  - ML et ML robuste: mêmes estimations que ML mais e.t. et chi-deux robustes à la non-normalité et à la non indépendance des observations (MLM, MLR)
  - WLS (données catégorielles),
  - ML-EM (données manquantes),
  - Bayes (MCMC)
- **Ajustement**
  - Chi-deux du rapport de vraisemblance, chi-deux robuste
  - AIC et BIC,
  - CFI et TLI,
  - RMSEA (*test of close fit*), SRMR (corrélations résiduelles moyennes),…

Stratégie de modélisation

- Description préalable des données
- Détermination de la forme de la courbe de croissance : examen des données moyennes et individuelles
- Etude du changement de la variance dans le temps
- Ajustement du modèle en fixant les occasions de mesure et sans les covariables
- Modifications éventuelles du modèle,
- Introduction des covariables

Introduction
GLM
MLR
LGC

Guide pratique

Soc Personal Psychol Compass. 2009 December 1; 3(6): 979–991. doi:10.1111/j.1751-9004.2009.00224.x.

The ABC's of LGM: An Introductory Guide to Latent Variable Growth Curve Modeling

Terry E. Duncan, Ph.D. and **Susan C. Duncan, Ph.D.**
Oregon Research Institute, 1715 Franklin Blvd. Eugene, OR 97403

Duncan, T. E., & Duncan, S. C. (2004). An introduction to latent growth curve modeling. *Behavior Therapy*, 35:2, 333-363.

http://dx.doi.org/10.1016/S0005-7894(04)80042-X

Introduction
GLM
MLR
LGC

Illustration : LGC avec Mplus

```
TITLE:  Modèle de croissance pour données continues
        VD: lect2 lect3 lect4 lect5 lect6;
        covariables invariantes dans le temps:
        int1 PISyl1 PIPH1 PESY1 PEPH1;
        covariables variant avec le temps:
        Let2 Let3 Let4 Let5 Let6;

MODEL:  !changement linéaire
        !i s | lect2@0 lect3@1 lect4@2 lect5@3 lect6@4;
        !changement quadratique
        i s q | lect2@0 lect3@1 lect4@2 lect5@3 lect6@4;

        !variance résiduelle de Y
        !lect2-lect6 (1);
        lect2@0;

        !var. et cov. des facteurs de croissance
        !i with s q;
        i with s q@0;
        s with q;
```

Illustration : LGC avec Mplus

Facteurs de croissance

		Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
Means					
	I	12.767	1.244	10.263	0.000
	S	4.473	0.541	8.261	0.000
	Q	-0.316	0.111	-2.841	0.004
Variances					
	I	92.845	16.951	5.477	0.000
	S	15.021	2.808	5.349	0.000
	Q	0.557	0.149	3.749	0.000
I WITH					
	S	-16.008	3.158	-5.069	0.000
	Q	0.000	0.000	999.000	999.000
S WITH					
	Q	-2.534	0.615	-4.121	0.000

Illustration : LGC avec Mplus

Vecteur de réponse

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
Intercepts				
LECT2	0.000	0.000	999.000	999.000
LECT3	0.000	0.000	999.000	999.000
LECT4	0.000	0.000	999.000	999.000
LECT5	0.000	0.000	999.000	999.000
LECT6	0.000	0.000	999.000	999.000
Residual Variances				
LECT2	0.000	0.000	999.000	999.000
LECT3	8.013	1.693	4.733	0.000
LECT4	5.090	1.570	3.242	0.001
LECT5	3.252	1.225	2.654	0.008
LECT6	1.492	2.506	0.595	0.552

Illustration : LGC avec Mplus

```
MODEL: i s q | lect2@0 lect3@1 lect4@2 lect5@3 lect6@4;
       lect2@0;
       i with s q@0;
       s with q;

       !introduction de cov. qui varient dans le temps (TVC)
       !réduction de la variance résiduelle de niveau 1?
       lect2 on Let2;
       lect3 on Let3;
       lect4 on Let4;
       lect5 on Let5;
       lect6 on Let6;

       i with let2-let6@0;
       s with let2-let6@0;
       q with let2-let6@0;

       let2 with let3-let6;
       let3 with let4-let6;
       let4 with let5-let6;
       let5 with let6;
```

Illustration : LGC avec Mplus

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
1) Sans Time-Varying Cov.				
Residual Variances				
LECT2	0.000	0.000	999.000	999.000
LECT3	8.013	1.693	4.733	0.000
LECT4	5.090	1.570	3.242	0.001
LECT5	3.252	1.225	2.654	0.008
LECT6	1.492	2.506	0.595	0.552
2) Avec Time-Varying Cov.				
Residual Variances				
LECT2	0.000	0.000	999.000	999.000
LECT3	7.546	1.669	4.522	0.000
LECT4	6.173	1.829	3.374	0.001
LECT5	3.675	1.353	2.716	0.007
LECT6	2.890	2.503	1.155	0.248

Illustration : LGC avec Mplus

```
MODEL:  i s q | lect2@0 lect3@1 lect4@2 lect5@3 lect6@4;
        lect2@0;
        i with s q@0;
        s with q;

        !introduction de cov. temporellement invariantes(TIC)
        !réduction de la variance résiduelle de niveau 2?
        !i s q on int1 PISyl1 PIPH1 PESY1 PEPH1;
         i on int1@0 PISyl1@0 PIPH1 PESY1@0 PEPH1;
         s on int1@0 PISyl1@0 PIPH1@0 PESY1@0 PEPH1;
         q on int1@0 PISyl1@0 PIPH1@0 PESY1@0 PEPH1@0;
```

Illustration : LGC avec Mplus

Facteurs de croissance on TIC

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
I ON				
INT1	0.000	0.000	999.000	999.000
PISYL1	0.000	0.000	999.000	999.000
PIPH1	0.300	0.351	0.853	0.393
PESY1	0.000	0.000	999.000	999.000
PEPH1	2.388	0.367	6.513	0.000
S ON				
INT1	0.000	0.000	999.000	999.000
PISYL1	0.000	0.000	999.000	999.000
PIPH1	0.000	0.000	999.000	999.000
PESY1	0.000	0.000	999.000	999.000
PEPH1	-0.377	0.083	-4.535	0.000
Q ON				
INT1	0.000	0.000	999.000	999.000
PISYL1	0.000	0.000	999.000	999.000
PIPH1	0.000	0.000	999.000	999.000
PESY1	0.000	0.000	999.000	999.000
PEPH1	0.000	0.000	999.000	999.000

Illustration : LGC avec Mplus

Facteurs de croissance

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
Intercepts				
I	5.157	1.579	3.267	0.001
S	5.530	0.575	9.618	0.000
Q	-0.317	0.111	-2.851	0.004
Variances residuelles				
I	50.172	9.421	5.326	0.000
S	14.125	2.778	5.084	0.000
Q	0.567	0.149	3.801	0.000
Covariances résiduelles				
I WITH				
S	-9.525	1.970	-4.835	0.000
Q	0.000	0.000	999.000	999.000
S WITH				
Q	-2.557	0.615	-4.155	0.000

Illustration : LGC avec Mplus

Vecteur de réponse

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
Intercepts				
LECT2	0.000	0.000	999.000	999.000
LECT3	0.000	0.000	999.000	999.000
LECT4	0.000	0.000	999.000	999.000
LECT5	0.000	0.000	999.000	999.000
LECT6	0.000	0.000	999.000	999.000
Residual Variances				
LECT2	0.000	0.000	999.000	999.000
LECT3	8.042	1.704	4.721	0.000
LECT4	4.984	1.561	3.192	0.001
LECT5	3.375	1.249	2.701	0.007
LECT6	1.077	2.566	0.420	0.675

Introduction
GLM
MLR
LGC

En résumé : MLR *vs* LGC

- **MLR** : les mesures du temps sont des **données** (statut de variable explicative).
  - Approche **univariée** ; la mesure du temps peut varier entre individus.
  - Effets **aléatoires** des covariables qui varient dans le temps.
  - **Niveaux supplémentaires** aisés à introduire.
- **SEM** : les mesures du temps sont des **paramètres**.
  - Approche **multivariée** ; la mesure du temps ne peut pas varier entre individus.
  - Effets **fixes** des covariables qui varient dans le temps.
  - Plus grande **flexibilité** (par ex., processus parallèles, VL prédicteurs de variables distales, etc.)