

PRINCIPES ET APPLICATIONS DE L'ANALYSE À NIVEAUX MULTIPLES EN ÉDUCATION

Gilles Raïche, UQAM

*Centre de soutien et d'animation en
évaluation des apprentissages à
l'enseignement supérieur*

28 mars 2012

CONTENU DE L'ATELIER

- Problématique
- Exemples qui seront utilisés
- Application dans R : librairie *lme4*
- Modèles à deux et trois niveaux : formulation
- Modélisation à deux niveaux
- Modèles de croissance
- Modélisation à trois niveaux
- Modélisations linéaires généralisées et non linéaires
- Logiciels disponibles et autres librairies de R
- Références

PROBLÉMATIQUE (1/3)

- **Modèles**

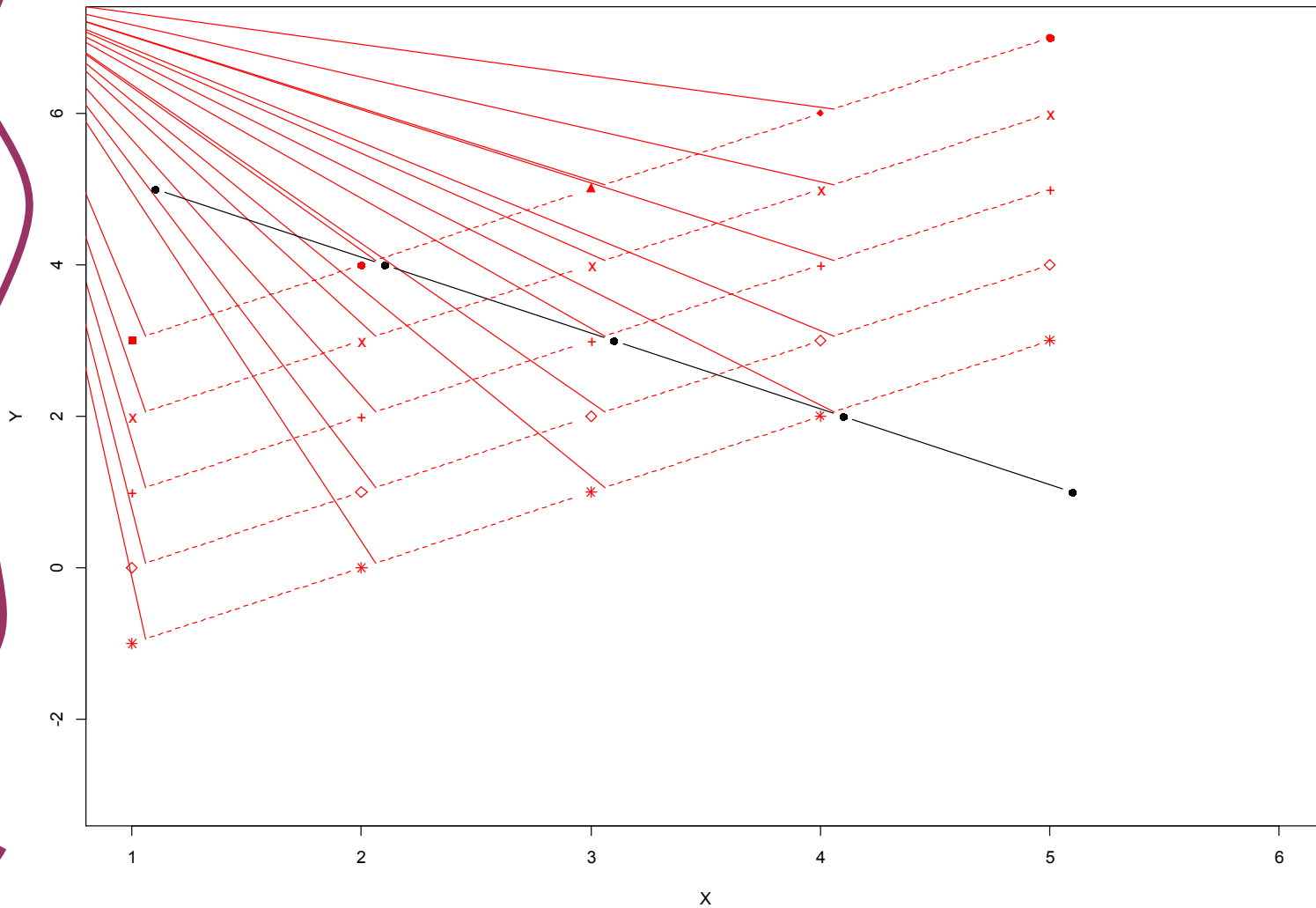
- Modèles mixtes : effets fixes et aléatoires
- Composantes de variance
- Modèles à coefficients aléatoires
- Modélisation à niveaux multiples
- Modélisations hiérarchiques : linéaires, linéaires généralisées et non linéaires

- **Applications**

- Éducation : structure à niveaux multiples
- Méta analyse
- Théorie de la généralisabilité

PROBLÉMATIQUE (2/3)

Erreur écologique (*ecological fallacy*)



PROBLÉMATIQUE (3/3)

- Meilleur calcul des erreurs types
- Augmentation de la puissance des tests
- Modélisations plus conforme à la réalité
- Toutefois, modélisations plus complexes

EXEMPLES QUI SERONT UTILISÉS (1/4)

- Étude sur le bien-être conditionnellement à la cohésion du groupe (Bliese et Halverson, 1996)
- Données disponibles dans la librairie *multilevel* de R
- Nombre de sujets : 7382
- Variables utilisées :
 - Sentiment de bien-être (`WBEING`)
 - Heures de travail individuel (`HRS`) et moyenne du groupe (`G.HRS`)
 - Considération individuelle du leadership (`LEAD`) et moyenne du groupe (`G.EAD`)
 - Groupe (`GRP`) : 1 à 99
 - Secteur -variable créée pour cet exemple (`SECTEUR`) : 0 à 10

EXEMPLES QUI SERONT UTILISÉS (2/4)

	<u>SECTEUR</u>	<u>GRP</u>	<u>LEAD</u>	<u>G.LEAD</u>	<u>HRS</u>	<u>G.HRS</u>	<u>WBEING</u>
1	1	1	3.18	2.93	12	10.97	2.11
2	1	1	3.00	2.93	11	10.97	3.33
3	1	1	3.64	2.93	12	10.97	2.11
4	1	1	3.36	2.93	9	10.97	4.39
5	1	1	3.55	2.93	7	10.97	1.72
6	1	1	3.25	2.93	8	10.97	3.33
7	1	1	3.64	2.93	9	10.97	3.44
8	1	1	3.27	2.93	8	10.97	3.67
9	1	1	3.55	2.93	12	10.97	3.06
10	1	1	2.55	2.93	11	10.97	1.72
11	1	1	3.45	2.93	11	10.97	3.67
12	1	1	4.00	2.93	10	10.97	4.28

EXEMPLES QUI SERONT UTILISÉS (3/4)

- Étude sur la trajectoire de la satisfaction au travail dans différentes entreprises (Bliese et Ployhart, 2002)
- Données disponibles dans la librairie *multilevel* de R
- Nombre de sujets : 1485
- Variables utilisées :
 - Satisfaction individuelle au travail mesurée à trois moments (SATISFACTION) : score de 1 à 5
 - Moment de la mesure de satisfaction (MOMENT) : 1 à 3
 - Sujet (SUBNUM) : 1 à 1485
 - Employeur (COMPANY) : A, B, C, D, F, HHC, REC et SVC

EXEMPLES QUI SERONT UTILISÉS (4/4)

	<u>COMPANY</u>	<u>SUBNUM</u>	<u>MOMENT</u>	<u>SATISFACTION¹</u>
1	HHC	1	0	1.666667
2	HHC	1	1	1.000000
3	HHC	1	2	3.000000
4	D	2	0	3.666667
5	D	2	1	4.000000
6	D	2	2	4.000000
7	B	3	0	4.000000
8	B	3	1	4.000000
9	B	3	2	4.000000
10	B	4	0	3.333333
11	B	4	1	3.666667
12	B	4	2	3.000000

Format de donnée long (et non pas large)

APPLICATION DANS *R* : LIBRAIRIE *LME4*

- Préparation des données dans un format adéquat: *long* et non *large* pour les mesures répétées
- Fonctions *lme()* et *gls()*

MODÉLISATIONS À DEUX ET TROIS NIVEAUX : FORMULATION (1/4)

Modèle nul (de référence ANOVA à coefficient aléatoire)

Niveau 1(sujets) :

$$WBEING = b_0 + e$$

Niveau 2 (groupes) :

$$b_0 = \beta_0 + u_0$$

MODÉLISATIONS À DEUX ET TROIS NIVEAUX : FORMULATION (2/4)

Modélisation à deux niveaux

Niveau 1(sujets) :

$$WBEING = b_0 + b_1HRS + b_2LEAD + e$$

Niveau 2 (groupes) :

$$b_0 = \beta_{00} + \beta_{01}G.HRS + \beta_{02}G.LEAD + u_{00}$$

$$b_1 = \beta_{10} + \beta_{11}G.HRS + \beta_{12}G.LEAD + u_{10}$$

$$b_2 = \beta_{20} + \beta_{21}G.HRS + \beta_{22}G.LEAD + u_{20}$$

MODÉLISATIONS À DEUX ET TROIS NIVEAUX : FORMULATION (3/4)

Modélisation à trois niveaux

Niveau 1 (sujets) :

$$WBEING = b_0 + b_1 HRS + b_2 LEAD + e$$

Niveau 2 (groupes) :

$$b_0 = \beta_{00} + \beta_{01} G.HRS + \beta_{02} G.LEAD + u_0$$

$$b_1 = \beta_{10} + \beta_{11} G.HRS + \beta_{12} G.LEAD + u_1$$

$$b_2 = \beta_{20} + \beta_{21} G.HRS + \beta_{22} G.LEAD + u_2$$

Niveau 3 (secteurs) :

$$\beta_{00} = \alpha_{000} + v_{00}$$

$$\beta_{01} = \alpha_{010} + v_{01}$$

$$\beta_{02} = \alpha_{020} + v_{02}$$

.....

$$\beta_{22} = \alpha_{220} + v_{22}$$

MODÉLISATIONS À DEUX ET TROIS NIVEAUX : FORMULATION (4/4)

Modèle de croissance (trajectoire) à deux niveaux

Niveau 1 (mesures) :

$$SATISFACTION = b_0 + b_1 MOMENT + e$$

Niveau 2 (sujets) :

$$b_0 = \beta_{00} + u_0$$

$$b_1 = \beta_{10} + u_1$$

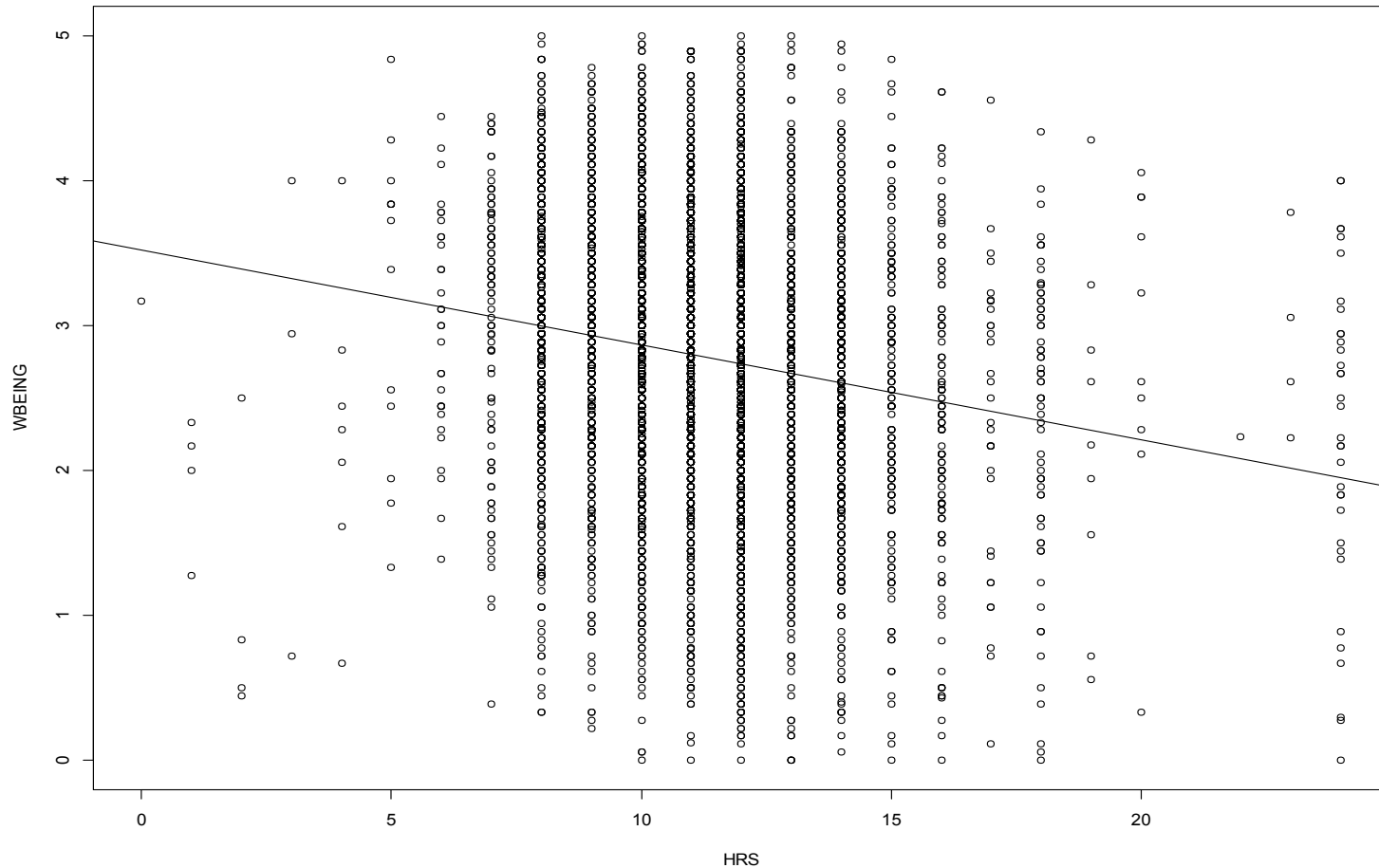
Niveau 3 (compagnies) :

$$\beta_{00} = \alpha_{000} + v_{00}$$

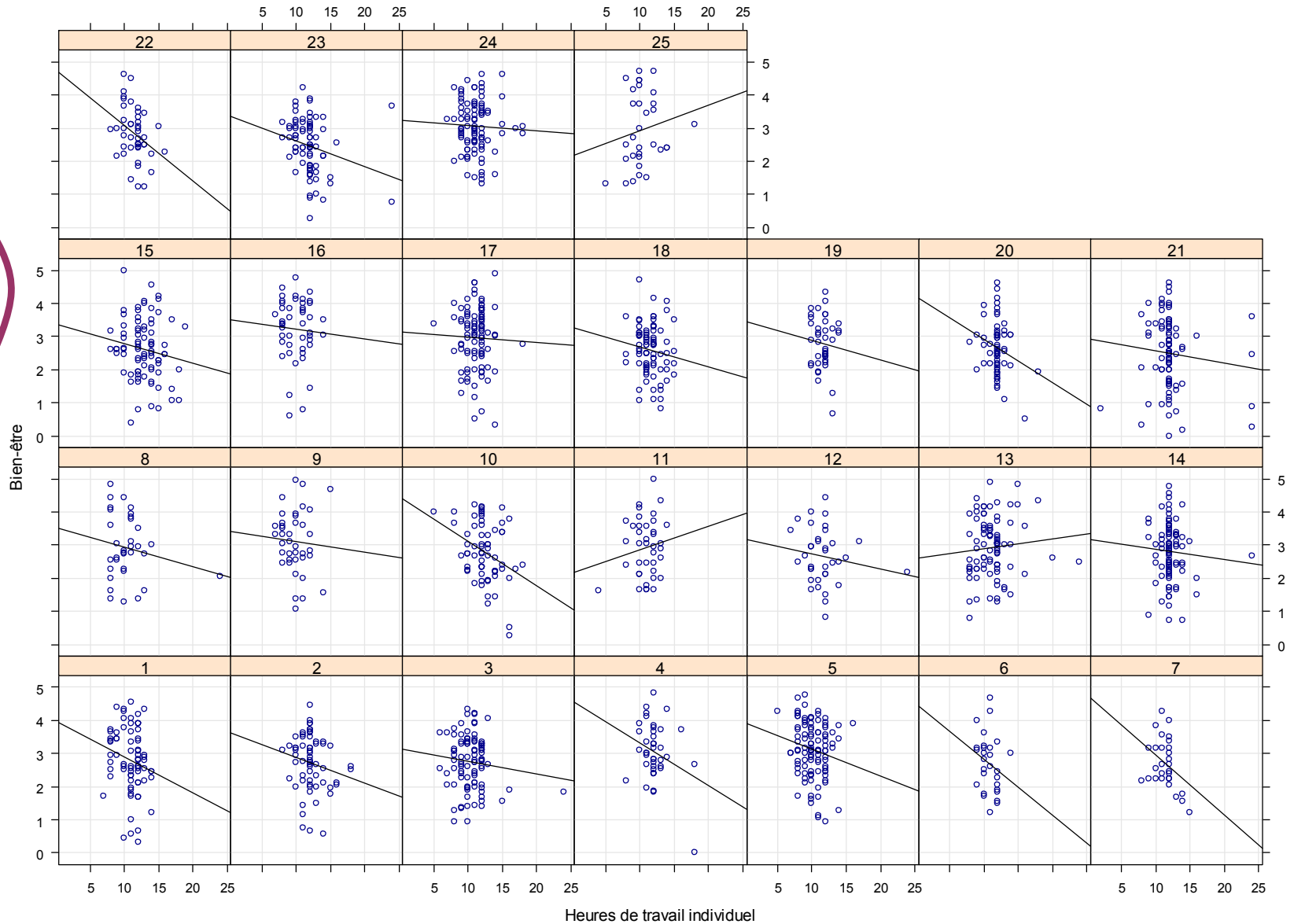
$$\beta_{01} = \alpha_{010} + v_{01}$$

MODÉLISATION À DEUX NIVEAUX (1/5)

- Nous allons étudier le bien-être au travail en fonction du nombre individuel d'heures travaillées



MODÉLISATION À DEUX NIVEAUX (2/5)



MODÉLISATION À DEUX NIVEAUX (3/5)

```
> # Modèle nul de comparaison
> model.00 <- lme(WBEING ~ 1, random=~1 | GRP, data=bh1996)
>
> model.02 <- lme(WBEING ~ HRS + G.HRS,
                  random=~1 + HRS | GRP, data=bh1996)
> anova(model.00, model.02)
```

	Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
model.00	1	3	19353.34	19374.06	-9673.669			
model.02	2	7	19223.81	19272.15	-9604.904	1 vs 2	137.5306	<.0001

MODÉLISATION À DEUX NIVEAUX (4/5)

```
> coef(model.02)
```

	<u>(Intercept)</u>	<u>HRS</u>	<u>G.HRS</u>
1	4.780577	-0.05172807	-0.1273416
2	4.839757	-0.05037962	-0.1273416
3	4.557961	-0.04020227	-0.1273416
4	5.006787	-0.05365535	-0.1273416
5	4.943264	-0.05611956	-0.1273416
6	4.688252	-0.04669054	-0.1273416
7	4.768233	-0.05049321	-0.1273416
8	4.761100	-0.04826678	-0.1273416
9	4.802504	-0.04848076	-0.1273416
10	5.014397	-0.05902430	-0.1273416
11	4.713763	-0.04124876	-0.1273416
12	4.715494	-0.04627961	-0.1273416

17	4.839406	-0.04646051	-0.1269261

MODÉLISATION À DEUX NIVEAUX (5/5)

```
> anova(model.02) # Coefficients fixes  
> VarCor(model.01) # Coefficients aléatoires
```

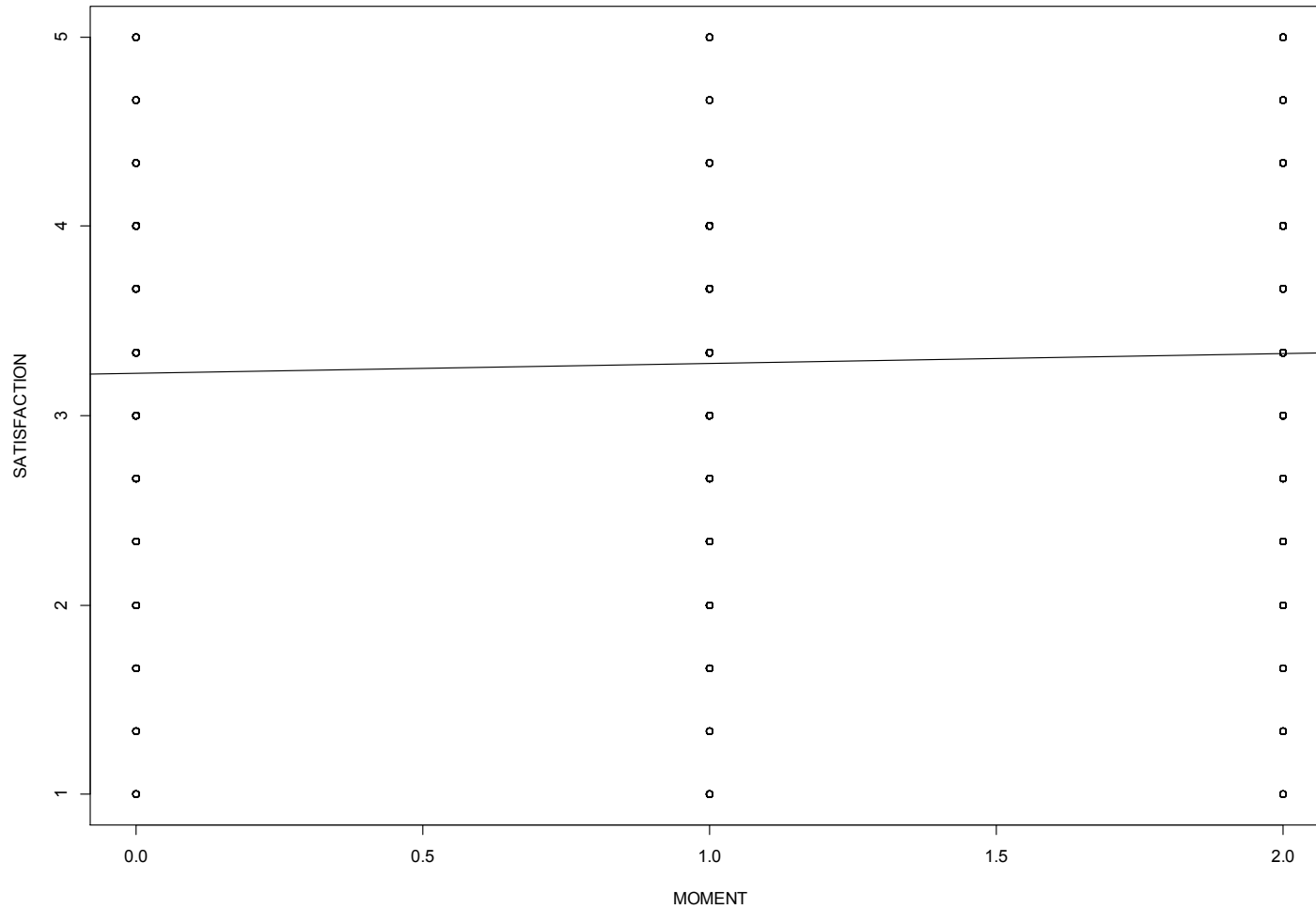
Coefficients fixes :

	<u>numDF</u>	<u>denDF</u>	<u>F-value</u>	<u>p-value</u>
(Intercept)	1	7282	28619.910	<.0001
HRS	1	7282	109.694	<.0001
G.HRS	1	97	42.056	<.0001

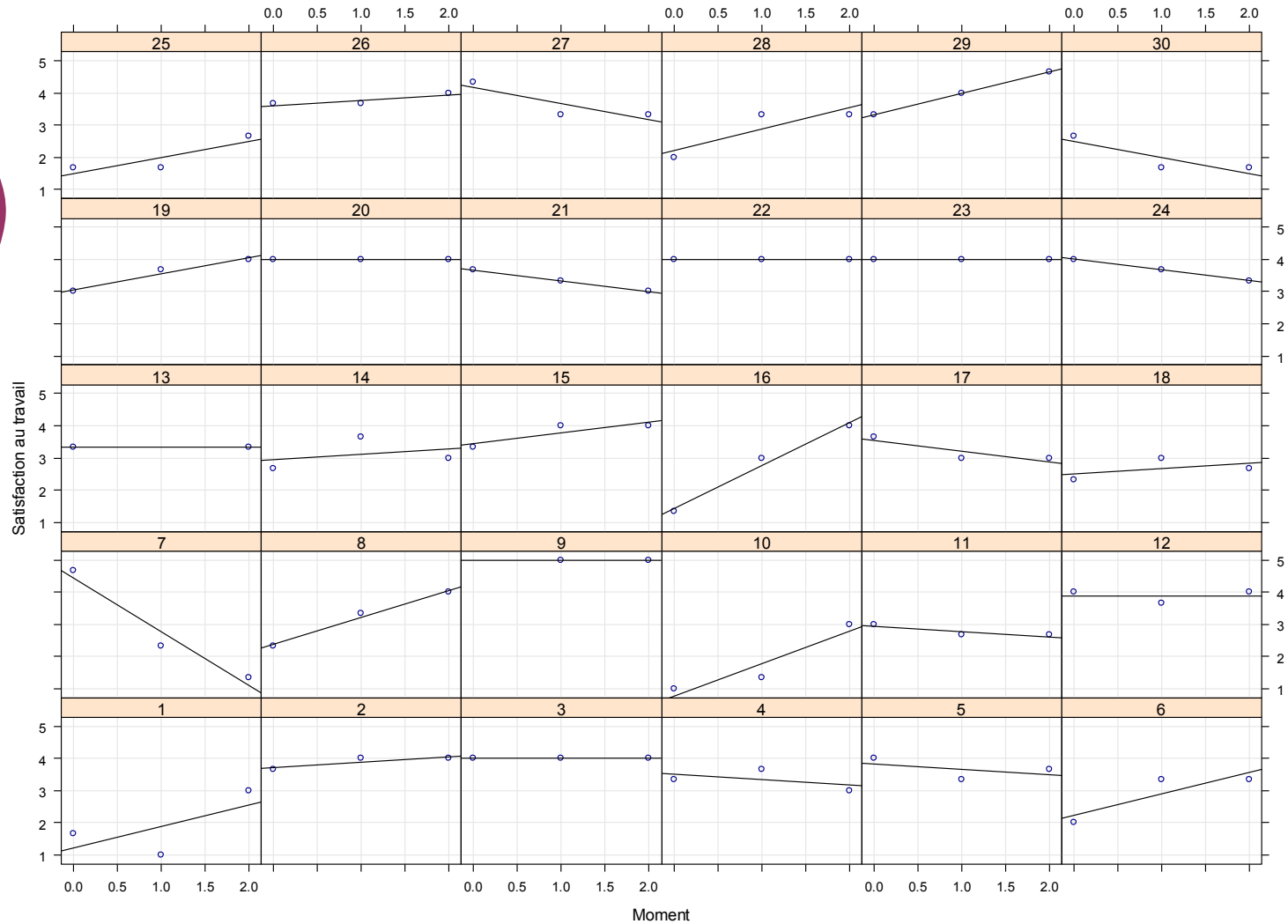
Coefficients aléatoires :

	<u>Variance</u>	<u>StdDev</u>	<u>Corr</u>
(Intercept)	0.0737602461	0.27158838	(Intr)
HRS	0.0003885055	0.01971054	-0.903
Residual	0.7780371663	0.88206415	

MODÈLES DE CROISSANCE (1/5)



MODÈLES DE CROISSANCE (2/5)



MODÈLES DE CROISSANCE (3/5)

```
> # Modèle nul de comparaison
> grow.00 <- lme(SATISFACTION ~ 1, random= ~ 1|SUBNUM,
                 data=grow.data, na.action=na.omit)
>
> grow.03 <- lme(SATISFACTION ~ 1 + MOMENT,
                 random= ~ 1 + MOMENT|SUBNUM,
                 data=grow.data, na.action=na.omit)
> anova(grow.00, grow03)
```

Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
grow.00	1	3454.209	3469.933	-1724.104			
grow.03	2	3423.603	3455.051	-1705.802	1 vs 2	36.60571	<.0001

MODÈLES DE CROISSANCE (4/5)

```
> coef(grow03)
```

	<u>(Intercept)</u>	<u>MOMENT</u>
1	1.778302	0.3540868894
2	3.703688	0.0619119794
3	3.866773	-0.0054306452
4	3.359663	-0.0307047533
5	3.646889	-0.0487341201
6	2.639981	0.2999988070
7	3.487037	-0.5526521141
8	2.859865	0.3433024659
9	4.468331	0.0478934046
10	1.547874	0.4827625578
11	2.880953	-0.0006558207
12	3.771031	0.0005791377

495	3.781575	-0.1713999876

MODÈLE DE CROISSANCE (5/5)

```
> anova(grow.03) # Coefficients fixes  
> VarCor(grow.03) # Coefficients aléatoires
```

Coefficients fixes :

	<u>numDF</u>	<u>denDF</u>	<u>F-value</u>	<u>p-value</u>
(Intercept)	1	903	9081.569	<.0001
MOMENT	1	903	5.078	0.0245

Coefficients aléatoires :

	<u>Variance</u>	<u>StdDev</u>	<u>Corr</u>
(Intercept)	0.66806720	0.8173538	(Intr)
MOMENT	0.09992208	0.3161045	-0.578
Residual	0.32865688	0.5732860	

MODÉLISATIONS À TROIS NIVEAUX (1/3)

```
> # Modèle nul de comparaison
> grow.00 <- lme(SATISFACTION ~ 1, random= ~ 1 | COMPANY/SUBNUM,
                data=grow.data, na.action=na.omit)
>
> grow.04 <- lme(SATISFACTION ~ 1 + MOMENT,
                random= ~ 1 + MOMENT | COMPANY/SUBNUM,
                data=grow.data, na.action=na.omit)
> anova(grow.00, grow04)
```

	Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
grow.004	1	4	3454.689	3475.655	-1723.345			
grow.04	2	9	3428.306	3475.478	-1705.153	1 vs 2	36.38366	<.0001

MODÉLISATIONS À TROIS NIVEAUX (2/3)

```
> coef(grow04)
```

	(Intercept)	MOMENT
A/8	2.858362	3.437252e-01
A/12	3.767017	7.556824e-04
A/13	3.285885	3.070652e-02
A/17	3.330217	-8.661521e-02
A/19	3.278986	2.029790e-01
A/22	3.862286	-5.174903e-03
A/24	3.711294	-1.103376e-01
A/30	2.349662	-8.285584e-02
A/35	3.436086	3.698970e-03
A/45	3.862286	-5.174903e-03
A/50	3.183617	-3.066132e-01

SVC/463	3.701081	1.766755e-02

MODÉLISATIONS À TROIS NIVEAUX (3/3)

```
> anova(grow.03) # Coefficients fixes  
> VarCor(grow.03) # Coefficients aléatoires
```

Coefficients fixes :

	<u>numDF</u>	<u>denDF</u>	<u>F-value</u>	<u>p-value</u>
(Intercept)	1	903	2403.4570	<.0001
MOMENT	1	903	5.1445	0.0236

Coefficients aléatoires :

	<u>Variance</u>	<u>StdDev</u>	<u>Corr</u>
COMPANY = pdLogChol(1 + MOMENT)			
(Intercept)	2.000116e-02	1.414255e-01	(Intr)
MOMENT	4.553555e-09	6.748003e-05	-0.002
SUBNUM = pdLogChol(1 + MOMENT)			
(Intercept)	6.541335e-01	8.087852e-01	(Intr)
MOMENT	1.000629e-01	3.163271e-01	-0.579
Residual	3.285301e-01	5.731755e-01	

MODÉLISATIONS LINÉAIRES GÉNÉRALISÉES ET NON LINÉAIRES

- Modélisations linéaires généralisées

$$y = \frac{1}{1 - e^{-b_0 + b_1 x_1}} + e$$

- Modélisations non linéaires

$$y = \sin(b_1 x_1) + \frac{1}{\cos(b_2 x_2)} + e$$

- Analyses factorielles confirmatoires et exploratoires

LOGICIELS DISPONIBLES ET LIBRAIRIES DE *R*

- *HLM (Bryk, Raudenbush et Congdon)*
- *MIWin (Goldstein)*
- *VarCL (Longford)*
- *SPSS (mixed)*
- *SAS (nlmixed)*
- *Systat*
- *Mplus (Muthén)*
- *R*
 - *lme4*
 - *multilevel*
 - *glmm*

RÉFÉRENCES

Bliese, P. D. et Halverson, R. R. (1996). Individual and nomothetic models of job stress : an examination of work hours, cohesion, and well-being. *Journal of applied social psychology*, 26, 1171-1189.

Bliese, P. D. et Ployhart, R. E. (2002). Growth modeling using random coefficient models: Model building, testing and illustrations. *Organizational research methods*, 5, 362-387.

Hox, J. (2002). *Multilevel analysis. Techniques and applications*. Mahwah, New Jersey : Lawrence Erlbaum Associates.

Raudenbush, S. W. et Bryk, A. S. (2002). *Hierarchical linear models. Applications and data analysis methods*, 2^e édition. Thousand Oaks, Californie : Sage.

Snijders, T. et Bosker, R. (1999). *Multilevel analysis. An introduction to basic and advanced multilevel modeling*. Thousand Oaks, Californie : Sage.

CONTACT

- Gilles Raîche
 - <http://www.Cdame.uqam.ca>
 - Raiche.Gilles@uqam.ca