

Introduction à l'analyse multiniveau

Hans Ivers

2 septembre 2020

Table des matières

Changements	2
1 Rappel sur la régression linéaire	3
2 Modèle multiniveau	3
2.1 Pourquoi cette technique?	3
2.2 L'approche longitudinale normale	4
3 Exemple SPSS	7
3.1 Régression linéaire	7
3.2 Justification du modèle multiniveau	9
3.2.1 Syntaxe SPSS	9
3.3 Modèle multiniveau normal	10
3.3.1 Examen visuel des trajectoires	10
3.3.2 Syntaxe SPSS	11
3.3.3 Effets fixes	12
3.3.4 Effets aléatoires	14
4 Modèle multiniveau généralisé	16
4.1 Introduction aux modèles binaire et ordinal	16
4.2 Exemple SAS pour modèle ordinal	18
Références	22

Changements

2019-05-12 :

- version initiale

2019-05-12 :

- explication de chaque mot clé pour les commandes SPSS
- ajout 2e conséquence si non-gestion de la dépendance des observations (section 2.1)
- ajout du modèle avec ordonnée aléatoire et calcul ICC (section 3.2)

2019-05-13 :

- ajout du modèle multiniveau généralisé et exemple SAS (section 4)

2020-09-02 :

- ajout de quelques références importantes

1 Rappel sur la régression linéaire

La régression linéaire a pour objectif de modéliser la relation entre une (ou plusieurs) variable indépendante X_s mesurée auprès des sujets $s = 1, \dots, N$, et une variable dépendante Y_s mesurée auprès de ces mêmes sujets. Le modèle prend la forme de :

$$Y_s = \beta_0 + \beta_1 X_s + \epsilon_s \quad (1)$$

Rappelons la définition de chaque élément dans cette équation :

- β_0 est l'ordonnée à l'origine (la valeur moyenne de Y quand $X = 0$)
- β_1 est la pente de la droite de régression, qui capture la relation entre le prédicteur X et la variable dépendante Y . La valeur de cette pente correspondant à la variation du nombre d'unités de Y pour chaque variation d'une unité de X (en conservant les autres X égaux s'il y a plus d'un prédicteur dans la régression)
- ϵ_s est l'erreur de prédiction, qu'on assume (1) indépendante entre les sujets, et (2) distribuée selon une loi normale de moyenne 0 et de variance σ^2 . Ces derniers postulats sont importants et nous y reviendrons plus loin.

Dans ce modèle, on dira que la portion $\beta_0 + \beta_1 X_s$ est la partie *fixe* de l'équation (ces coefficients β_0, β_1 sont des constantes une fois qu'ils sont estimés), alors que ϵ_s est la partie *aléatoire* (on ne connaît pas sa valeur exacte, seulement sa loi).

Enfin, on peut souligner que l'estimation des coefficients de régression est réalisée à l'aide de la méthode des moindres carrés, où l'objectif est de trouver les β_0, β_1 qui *minimisent* l'erreur quadratique (au carré) moyenne (EQM) de prédiction :

$$EQM = \frac{1}{N} \sum_{s=1}^N (Y_s - \hat{Y}_s)^2 \quad (2)$$

Sachant que la valeur prédite de la variable dépendante est $\hat{Y}_s = \hat{\beta}_0 + \hat{\beta}_1 X_s$, on peut voir ici que les meilleures estimations des coefficients sont celles qui permettent le plus petit écart entre la valeur observée Y_s et la valeur prédite \hat{Y}_s pour tous les $s = 1, \dots, N$ sujets, ce qui donne l'EQM la plus petite et maximise l'ajustement de la régression entre X et Y .

2 Modèle multiniveau

2.1 Pourquoi cette technique ?

Le modèle multiniveau (MM) est une approche plus générale de la régression, où on relaxe un des postulats, soit l'indépendance des observations [2, 5, 6, 7]. Ce modèle n'est pas naturellement longitudinal, il permet simplement de gérer plus efficacement la *dépendance entre les observations*. Le cas classique du MM est illustré dans un devis où on désire comparer des groupes d'élèves (ou étudier la relation entre des caractéristiques de ces élèves). On sait que ces élèves se regroupent dans des classes, qui partagent un même milieu de vie, un même professeur, etc. On peut donc postuler qu'il existe une certaine corrélation

entre les réponses (Y_s) des enfants provenant d'une même classe (une notion connue sous le nom de *corrélation intra-classe* ou ICC en anglais). Or, la régression linéaire ne peut pas gérer cette dépendance. En fait, toutes les analyses classiques vues dans les cours de statistiques de base (p.ex., test t, corrélation, régression, ANOVA, khi-carré) assument que les observations sont parfaitement indépendantes.

Pourquoi est-ce un problème de ne pas tenir compte de cette dépendance ? L'explication est en deux volets : (1) moins certaines observations sont indépendantes (plus il y a de corrélation entre elles, $ICC > 0$) et moins chaque observation apporte de l'information nouvelle. A la limite, si nous observons une corrélation parfaite ($r = 1$) entre les réponses des n élèves d'une même classe, une seule observation sera suffisante pour "résumer" le phénomène mesuré Y_s dans cette classe. Donc, le fait de ne pas tenir compte de la dépendance nous amène à surestimer la quantité d'information indépendante (le concept de degré de liberté) et donc à surestimer la signification des tests (car on sait qu'un grand N va naturellement réduire la p-valeur de n'importe quel test). (2) des observations corrélées auront moins tendance à varier librement, ce qui va contribuer à une sous-estimation de la variance réelle des observations, et ainsi produire des erreurs standards trop petites (et donc encore ici une surestimation de la signification des tests sur les paramètres).

En conclusion, la première raison pour laquelle on fait un MM est l'objectif d'avoir des tests statistiques valides [2], et non artificiellement significatifs car nous n'aurions pas tenu compte de la possible dépendance entre les observations à l'intérieur de chaque classe.

2.2 L'approche longitudinale normale

Dans l'approche longitudinale qui nous intéresse davantage, la structure multiniveau sera différente car l'origine de la corrélation entre les observations est différente. Dans l'exemple de l'école, la dépendance se situe entre les élèves d'une même classe alors dans un devis longitudinal, la dépendance sera plutôt observée entre les différentes mesures [2, 6] pour un même sujet¹. Ainsi, pour une hiérarchie à deux niveaux, la variable dépendante Y_{st} aura deux indices pour identifier clairement chaque niveau. Le premier indice s correspond au niveau le plus élevé (niveau 2), soit le sujet $s = 1, \dots, N$, alors que le second indice t correspond au niveau le plus bas (niveau 1), soit la mesure $t = 1, \dots, T$ réalisée auprès de ce s^e sujet, pour $T =$ nombre de temps de mesure.

Comment le modèle permet de "capturer" ce qui est commun à chaque unité de niveau 2 (classe d'élèves pour le devis transversal, ou chaque sujet pour le devis longitudinal) ? L'idée toute simple est de permettre que les paramètres de la régression puissent *varier d'une unité de niveau 2 à l'autre* car si toutes les observations d'un même unité de niveau 2 ont la même "équation de régression", ceci implique clairement qu'elles partagent quelque chose, qui est distinct de la régression pour une autre unité de niveau 2.

Le modèle de régression multiniveau prendra donc la forme suivante :

$$Y_{st} = \beta_{0s} + \beta_{1s}X_{st} + \epsilon_{st} \quad (3)$$

1. Notons qu'il est très possible d'avoir une structure hiérarchique dans un devis longitudinal, par exemple, des élèves d'une même classe qui sont chacun évalués à plusieurs reprises

Arrêtons-nous ici pour examiner *deux* implications de cette notation. Premièrement, comme les coefficients relatifs à l'ordonnée à l'origine (β_0) et à la pente (β_1 la contribution du prédicteur X) ont un indice $s = 1, \dots, N$, ceci implique que l'analyse va maintenant estimer ces deux coefficients pour *chaque* unité de niveau 2 (chaque sujet si devis longitudinal). Or, si l'analyse estime N coefficients pour l'ordonnée à l'origine et N coefficients pour la pente, elle peut donc estimer une moyenne et une variance pour la distribution de chaque paramètre (ordonnée ou pente). Un paramètre qui a une moyenne ET une variance est qualifié d'effet *aléatoire*[2], qu'on peut contraster avec l'équation (1) où chaque paramètre pouvait prendre une seule valeur pour tout l'échantillon, un comportement qualifié d'effet *fixe*.

Deuxièmement, cette notation permet d'introduire l'idée plus générale du modèle multi-niveau, soit que les paramètres de niveau inférieur sont estimés par des "hyper-paramètres" de niveau supérieur. En effet, on peut redéfinir le modèle multiniveau en couches, selon la notation suivante :

$$\text{(niveau 1)} \quad Y_{st} = \beta_{0s} + \beta_{1s}X_{st} + \epsilon_{st} \quad (4)$$

$$\text{(niveau 2 pour l'ordonnée)} \quad \beta_{0s} = \beta_0 + u_{0s} \quad (5)$$

$$\text{(niveau 2 pour la pente)} \quad \beta_{1s} = \beta_1 + u_{1s} \quad (6)$$

Selon cette approche, les deux paramètres aléatoires (β_{0s} et β_{1s}) du niveau 1 sont en fait des paramètres eux-mêmes estimés à l'aide d'autres paramètres plus "généraux" (d'où le nom "hyper-paramètre"). L'équation (5) définit la valeur de l'ordonnée aléatoire β_{0s} comme l'addition de l'ordonnée générale β_0 , qui ne change pas selon le sujet, et d'une certaine variation aléatoire u_{0s} , qui varie selon le sujet s . On peut constater la même définition à l'équation (6) pour la pente aléatoire. Est-ce que la valeur des β_0 et β_1 du MM seront les mêmes que ceux de la régression linéaire? C'est une question importante, que nous traiterons dans l'exemple SPSS.

L'intérêt de la définition par niveau (équations 4 à 6) est qu'on y distingue mieux la partie fixe et aléatoire. Il est possible de tracer ici un parallèle avec la régression linéaire (l'équation 1), où les β du MM correspondent à la partie fixe $\beta_0 + \beta_1 X$ de la régression linéaire, et où le u du MM correspond à la partie aléatoire σ^2 de la régression. En conséquence, comme pour l'erreur ϵ_s de la régression linéaire, on peut poser que la partie aléatoire u_s du MM suivra une loi normale, de moyenne 0 mais de variance τ^2 , *distincte* de la variance σ^2 des erreurs du modèle.

En résumé, on peut donc poser les postulats suivants sur les deux effets aléatoires du MM[9] :

1. la distribution de l'ordonnée à l'origine β_{0s} suit une loi normale de moyenne β_0 et de variance τ_0^2 , distincte de la variance σ^2 des erreurs du modèle.
2. On peut poser le même postulat pour la distribution de la pente β_{1s} , qui suit également une loi normale mais de moyenne β_1 et de variance τ_1^2 , distincte de la variance de l'ordonnée et des erreurs du modèle.
3. Si les ordonnées et les pentes peuvent varier librement entre les unités de niveau 2, il est possible que la variation de ces deux paramètres ne soit pas indépendante. Par

exemple, on peut penser que les sujets qui présentent une faible valeur de la variable dépendante en début de trajectoire (faible ordonnée à l'origine) auront davantage une trajectoire ascendante (pente positive) et vice-versa, ce qui se traduit statistiquement par une possible corrélation *négative* entre ces deux paramètres. On peut donc estimer la covariance τ_{01} entre les N ordonnées et N pentes.

En clair, chaque effet aléatoire dans un modèle multiniveau est associé à l'estimation d'un effet moyen (ici β_0 et β_1) et la variance de cet effet (τ_0^2 et τ_1^2), en plus de la possible covariance entre les effets aléatoires (τ_{01}). On dira alors que ces effets aléatoires suivent une loi normale bivariée \mathcal{N}_2 de moyennes (vecteur de 2 moyennes) et de variance-covariance (matrice 2×2) :

$$\begin{bmatrix} \beta_{0s} \\ \beta_{1s} \end{bmatrix} \sim \mathcal{N}_2 \left(\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, \begin{bmatrix} \tau_0^2 & \tau_{01} \\ \tau_{01} & \tau_1^2 \end{bmatrix} \right) \quad (7)$$

dont la forme particulière de la matrice de variance-covariance sera qualifiée de "non-structurée" (UNstructured en SAS[6] ou SPSS[8]).

Dans un devis longitudinal, la première approche statistique introduite est habituellement l'analyse de variance à mesures répétées. Cette analyse est connue sous le nom d'approche "marginale" (*population-averaged* en anglais) car on s'intéresse ici aux moyennes observées à chaque temps de mesure. Le temps est donc vu comme une variable nominale (différents niveaux d'un facteur "temps"), et on n'utilise pas l'information sur le moment réel de chaque mesure, seulement l'ordre dans lesquelles les mesures sont prises.

Cet usage limité de la variable temps amène son lot de conséquences négatives :

- l'information sur le moment exact de la mesure n'est pas utilisée dans le modèle
- on impose d'avoir un même nombre de mesures pour chaque sujet
- une mesure manquante devient automatiquement une donnée manquante, qui doit être gérée dans une étape ultérieure

Pour sa part, l'approche multiniveau est qualifiée de "basée sur les sujets" (*subject-specific* en anglais) car elle vise à modéliser la trajectoire temporelle de *chaque* sujet (i.e., unité de niveau 2). De plus, l'équation (3) permet de constater que le temps est introduit comme une variable continue dans un modèle multiniveau, ce qui amène son lot d'avantages[6]. En particulier, la notion de "temps manquant" n'est plus pertinente car on peut avoir un nombre variable de mesures, prises à des intervalles distincts selon le sujet. Le temps est donc traité de façon beaucoup plus flexible car l'intérêt de cette analyse est d'estimer la relation entre le temps X et la variable dépendante Y , comme une trajectoire de forme linéaire².

Toutefois, mentionnons deux principaux désavantages, qui n'en sont pas vraiment en pratique. Premièrement, il n'est plus possible d'obtenir (par défaut) une moyenne pour

2. Dans l'équation 1, le prédicteur $X = temps$ est linéaire donc la trajectoire estimée sera une droite linéaire. Toutefois, il est possible de modéliser le temps comme une trajectoire courbe, par exemple quadratique, en ajoutant simplement un terme qui capture le carré de l'effet du temps, ce qui donne le modèle polynomial suivant : $Y_{st} = \beta_{0s} + \beta_{1s}X_{st} + \beta_{2s}X_{st}^2 + \epsilon_{st}$

chaque temps de mesure³ car l'effet du temps est maintenant résumé par le coefficient de régression β_1 . Deuxièmement, le fait d'utiliser le temps en continu peut apporter certaines difficultés d'interprétation selon la nature de la mesure du temps (temps de calendrier vs. temps depuis un événement initial), mais ce défi peut être résolu par une réflexion sur la nécessité ou non de *centrer* cette variable.

3 Exemple SPSS

Pour illustrer l'estimation d'une régression linéaire vs. modèle multiniveau normal, un petit jeu de données a été généré pour $N = 5$ sujets avec $T = 3$ mesures chacun, pour un total de 15 observations.

Jeu de données		
id	x	y
1	0	10
1	1	12
1	4	14
2	2	9
2	3	10
2	5	12
3	1	6
3	3	7
3	5	8
4	0	12
4	1	11
4	3	16
5	2	9
5	3	11
5	5	14

3.1 Régression linéaire

On peut d'abord réaliser une régression linéaire simple pour prédire Y selon le temps X (en années). Cette analyse révèle que les estimations des coefficients sont respectivement de $\hat{\beta}_0 = 9.815$ et $\hat{\beta}_1 = 0.363$ (voir Figure 1). En clair, la trajectoire moyenne au temps $X = 0$ est de 9.815 et cette moyenne augmente de 0.363 pour chaque année qui passe, $t(ddf = 13) = 0.85, p = .41$. La p-valeur observée suggère que cette augmentation ne serait pas significativement différente de zéro (non-rejet de l'hypothèse nulle $H_0 : \beta_1 = 0$).

3. Il est exagéré de qualifier ce constat de limite car il est toujours possible d'obtenir une projection statistique de la moyenne des Y pour un $X =$ [une certaine valeur de temps] mais cette information n'est pas disponible par défaut dans les logiciels.

Coefficients^a

Modèle	Coefficients non standardisés		Coefficients standardisés	t	Sig.
	B	Erreur standard	Bêta		
1 (Constante)	9,815	1,299		7,558	,000
Temps	,363	,428	,229	,847	,412

a. Variable dépendante : Variable dépendante

FIGURE 1 – Régression pour prédire Y selon le temps X

L'erreur quadratique moyenne de ce modèle de régression est de $\hat{\sigma}^2 = 7.650$ (valeur indiquée comme le carré moyen du terme résiduel dans la table d'ANOVA de la régression).

Une illustration visuelle de la droite de régression estimée est présentée à la Figure 2. On peut constater que cette droite a pour objectif de résumer l'association générale entre X et Y pour tous les points, indépendamment du sujet. On voit bien ici l'idée de l'approche dite "marginale".

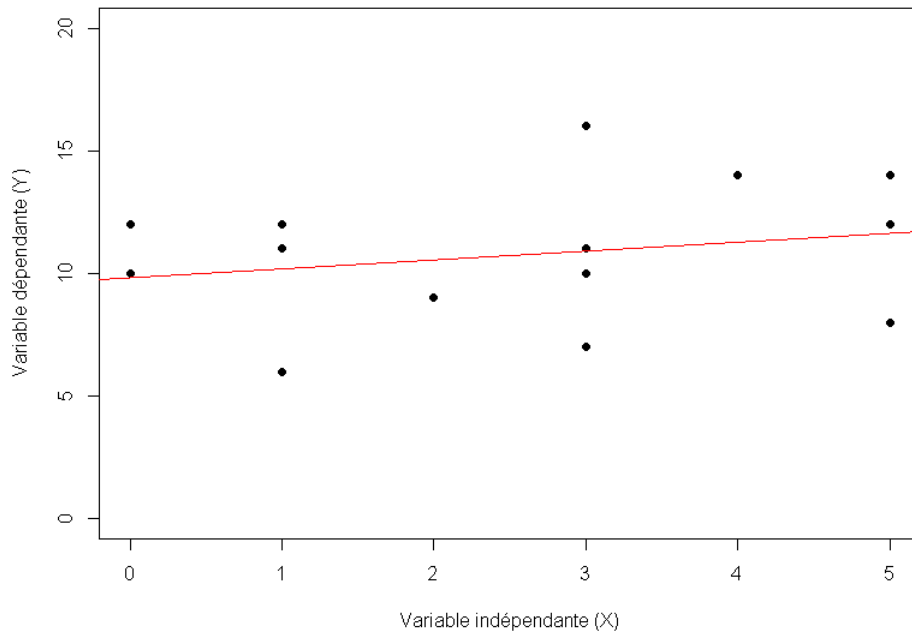


FIGURE 2 – Relation marginale prédite pour la régression linéaire

3.2 Justification du modèle multiniveau

Avant de réaliser un modèle multiniveau "complet", il est légitime de poser la question à savoir si ce modèle est nécessaire. En clair, est-ce qu'il existe des évidences d'une hétérogénéité entre les sujets qui pourrait justifier de réaliser un modèle multiniveau ? Car dans le cas contraire, une approche marginale sera suffisante.

Pour se faire, on réalise un premier modèle multiniveau très simple, contenant uniquement une ordonnée à l'origine β_0 aléatoire :

$$Y_{st} = \beta_{0s} + \epsilon_{st} \quad (8)$$

Ce type de modèle est connu sous le nom de *intercept-only model* [6, 9]. Comme l'effet du temps (la variable indépendante X) n'est pas présent dans ce modèle, l'intérêt est d'étudier si la variabilité des moyennes de chaque sujet (car l'ordonnée d'un modèle sans prédicteur équivaut à la moyenne des observations pour chaque sujet) est non-nulle.

3.2.1 Syntaxe SPSS

La syntaxe SPSS pour estimer ce modèle est présentée dans le tableau suivant :

Syntaxe SPSS pour MM avec ordonnée aléatoire	
MIXED y	
/PRINT=SOLUTION TESTCOV	
/RANDOM=INTERCEPT SUBJECT(id).	

A la première ligne (obligatoire), la procédure SPSS MIXED [8] spécifie la variable dépendante Y mais il n'y aucune variable indépendante dans ce modèle.

La ligne optionnelle /PRINT permet l'affichage de différentes informations, telles que la valeur finale des paramètres fixes (β_0) et aléatoires (σ^2 et τ_0^2), option SOLUTION ; et le test de Wald et l'intervalle de confiance à 95% pour l'effet aléatoire, option TESTCOV.

Enfin la ligne obligatoire /RANDOM permet de spécifier le seul effet aléatoire (INTERCEPT = ordonnée à l'origine), alors que l'option SUBJECT(id) précise la variable (id) qui indique les unités de niveau 2.

Seule la sortie sur les effets aléatoires (voir Figure 3) nous intéresse ici.

Estimations des paramètres de covariance ^a						
Paramètre	Estimation	Erreur standard	Z de Wald	Sig.	Intervalle de confiance à 95 %	
					Borne inférieure	Borne supérieure
Résidu	4,133333	1,848483	2,236	,025	1,720408	9,930459
Constante [sujet = id] Variance	3,922222	3,797980	1,033	,302	,587887	26,168012

a. Variable dépendante : Variable dépendante.

FIGURE 3 – Effets aléatoires du modèle avec ordonnée aléatoire

La ligne intitulée "Résidu" indique la variance estimée des résidus $Y_{st} - \hat{Y}_{st}$, soit $\hat{\sigma}_\epsilon^2 = 4.133$ (voir Equation 3) alors que la seconde ligne, plus ou moins clairement identifiée, indique en fait la variance de l'ordonnée aléatoire, $\hat{\tau}_0^2 = 3.922$. Est-ce que cette variance est "significative" ? Le test de Wald et sa signification, rapportés dans les 4e et 5e colonnes, peut être utilisé pour vérifier si cette variance est "significative". Toutefois, c'est un test généralement invalide car estimé sur la marge de la distribution (une variance ne pouvant pas être négative). C'est pour cela qu'on lui préfère l'examen de l'intervalle de confiance. Ici, les valeurs probables (à 95%) de cette variance se situent entre 0.59 et 26.17, ce qui exclut la valeur 0. On peut donc conclure qu'il existe une hétérogénéité significative entre les sujets, justifiant une approche multiniveau.

On peut même quantifier plus précisément l'hétérogénéité en calculant le coefficient de corrélation intra-classe (ICC), comme le ratio de la variance de l'ordonnée sur la somme des deux variances :

$$ICC = \frac{\tau_0^2}{\tau_0^2 + \sigma_\epsilon^2} = \frac{3.922}{3.922 + 4.133} = 0.487 \quad (9)$$

Selon cette définition mathématique, on peut donc affirmer que 48.7% de la variance de Y est attribuable aux différences entre les $N = 5$ sujets (pour être plus précis, entre les 5 moyennes par sujet). Un second constat, moins intuitif (voir [6] pour une discussion sur ce point), permet d'interpréter l'ICC comme un indice de corrélation entre les observations d'une même unité de niveau 2, ce qui suggère que les observations d'un même sujet présenteraient une assez forte corrélation entre elles.

3.3 Modèle multiniveau normal

Réalisons maintenant un modèle multiniveau "complet", où l'ordonnée à l'origine et la pente sont fixées comme effets aléatoires, ce qui équivaut à estimer une droite de régression (i.e., une trajectoire) par unité de niveau 2 (par sujet pour un devis longitudinal).

3.3.1 Examen visuel des trajectoires

Tout d'abord, examinons les trajectoires observées (en noir) et prédites (en bleu) à la Figure 4. On peut constater que chaque trajectoire prédite présente une pente plus abrupte (coefficient β_1 plus grand) que la pente marginale prédite par la modèle de régression linéaire (en rouge). Comment est-ce possible ? Simplement parce que la relation entre X et Y a été en quelque sorte "diluée" dans la régression linéaire car estimée pour l'ensemble des données, sans tenir compte du sujet. Cet important constat amène deux implications : (1) les coefficients des effets fixes (les β) estimés dans un MM sont généralement plus larges que ceux des modèles marginaux, mais (2) comme la valeur de ces effets est *conditionnelle* aux effets aléatoires (i.e., aux particularités des sujets), on dira que les conclusions sont restreintes à notre échantillon (*subject-specific*), comparativement aux conclusions sur les effets marginaux, qui n'ont pas cette restriction [1].

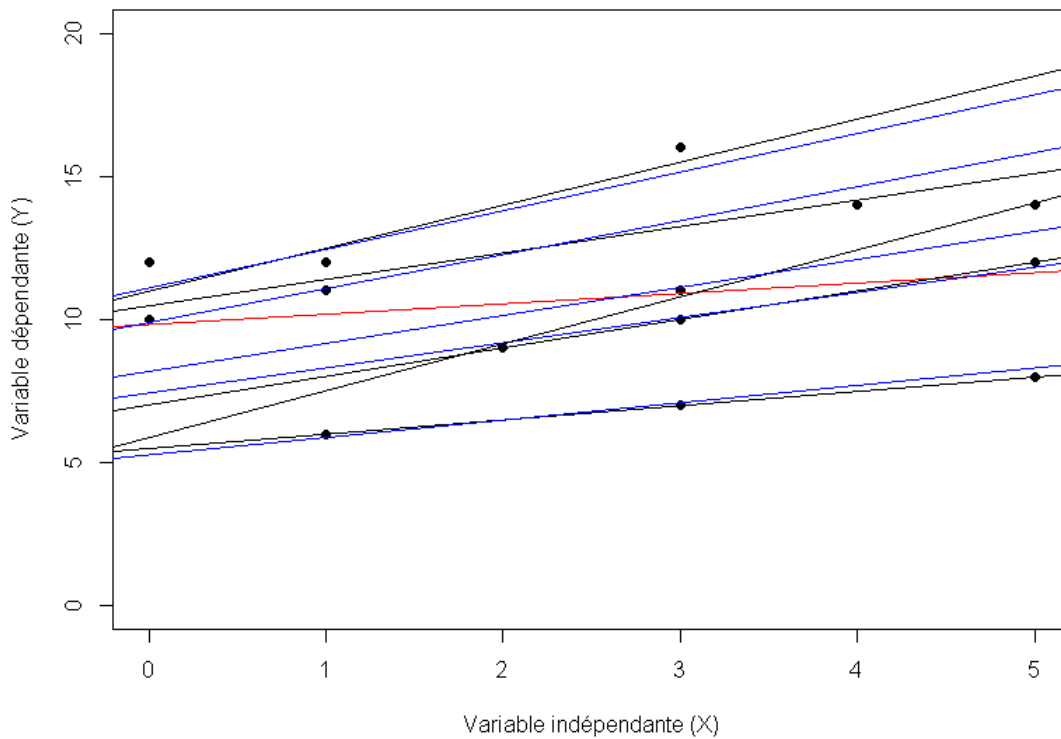


FIGURE 4 – Trajectoires individuelles observées (en noir) et prédites (en bleu)

Est-ce que les résultats de ces deux approches peuvent coïncider ? Si les sujets avaient tous un profil similaire, la pente marginale estimée par la régression linéaire sera très similaire à la moyenne des pentes par sujet estimées grâce au MM. Mais cette situation est rarement observée en pratique.. car si les sujets sont similaires, ils sont interchangeables et donc il n'y a plus lieu de réaliser une régression par sujet (i.e., le MM n'est plus nécessaire..).

3.3.2 Syntaxe SPSS

La syntaxe pour estimer ce modèle, connu sous le nom de *random coefficients model*[2] dans la littérature statistique, est présentée dans le tableau suivant :

Syntaxe SPSS pour MM avec ordonnée et pente aléatoires

```
MIXED y WITH x
/CRITERIA=MXITER(500) MXSTEP(100)
/FIXED=x
/METHOD=REML
/PRINT=HISTORY(5) SOLUTION TESTCOV
/RANDOM=INTERCEPT x | SUBJECT(id) COVTYPE(UN).
```

L'analyse MM est réalisée à l'aide de la procédure SPSS MIXED [8]. A la première ligne obligatoire, on indique d'abord la variable dépendante Y puis la variable indépendante X, précédée du mot clé WITH. Ce mot-clé est très important car il définit la variable indépendante comme une variable continue, alors que pour une ANOVA / modèle mixte habituel, on utilisera plutôt le mot-clé BY (qui définit la variable indépendante comme une variable nominale, en facteur).

La ligne optionnelle /CRITERIA permet d'augmenter le nombre par défaut d'itérations pour obtenir une solution convergente. Il est rarement nécessaire de changer les options par défaut en pratique. La ligne /FIXED= permet de spécifier la variable indépendante (seulement la variable X ici) alors que la ligne /METHOD indique la méthode d'estimation par défaut (REML pour REstricted Maximum Likelihood).

La ligne optionnelle /PRINT permet l'affichage de différentes informations, telles que la valeur des différents paramètres après chaque 5 itérations, HISTORY(5); la valeur finale des paramètres fixes et aléatoires, SOLUTION; et le test de Wald et l'intervalle de confiance à 95% pour les effets aléatoires, TESTCOV.

Enfin la ligne obligatoire /RANDOM permet de spécifier les deux effets aléatoires (INTERCEPT = ordonnée à l'origine et x pour l'indicateur de temps), alors que l'option SUBJECT(id) précise la variable (id) qui indique les unités de niveau 2. Enfin l'option COVTYPE(UN) indique que la matrice de variance-covariance des effets aléatoires est de type non-structurée (UNstructured, voir équation 7 pour une explication détaillée).

3.3.3 Effets fixes

Une série de tableaux est produite en sortie par SPSS. D'abord, le tableau des dimensions du modèle (voir Figure 5) permet de s'assurer que le modèle MM a été bien spécifié. On peut voir que $1 + 1 = 2$ paramètres sont à estimer pour les effets fixes β_0 et β_1 . Il y a trois paramètres pour les effets aléatoires (les deux variances τ_0^2, τ_1^2 , et leur covariance τ_{01}) et le dernier paramètre pour la variance de l'erreur σ^2 , pour un total de 6 paramètres.

La seconde figure de cette section présente l'estimation des paramètres des effets fixes (voir Figure 6). Ces résultats révèlent qu'au temps $X = 0$, une "trajectoire typique" d'un sujet débute avec un score moyen de $\hat{\beta}_0 = 8.38$ et augmente de $\hat{\beta}_1 = 1.00$ unité par année, $t(ddl = 1.86) = 4.70, p = .05$. On peut noter ici que, selon le seuil de $\alpha = 5\%$, la p-valeur observée permet le rejet de l'hypothèse nulle $H_0 : \beta_1 = 0$, ce qui supporte la conclusion que cette augmentation serait statistiquement significative.

Il y a plusieurs commentaires à faire sur ces résultats :

		Nombre de niveaux	Structure de covariance	Nombre de paramètres	Variables sujet
Effets fixes	Constante	1	Sans structure	1	id
	x	1		1	
Effets aléatoires	Constante + x	2		3	
Résidu				1	
Total		4		6	

a. Variable dépendante : Variable dépendante.

FIGURE 5 – Modèle MM - dimensions du modèle

Paramètre	Estimation	Erreur standard	ddl	t	Sig.	Intervalle de confiance à 95 %	
						Borne inférieure	Borne supérieure
Constante	8,376866	1,146873	3,668	7,304	,003	5,075651	11,678081
x	1,000320	,212788	1,858	4,701	,049	,014408	1,986232

a. Variable dépendante : Variable dépendante.

FIGURE 6 – Modèle MM - effets fixes estimés

1) Que veut-on dire par "trajectoire typique" ? En fait, les estimés de l'ordonnée et de la pente sont simplement la moyenne des estimés prédits pour les N trajectoires individuelles. Pourquoi ? Il faut se souvenir que chaque effet aléatoire dans le modèle MM est distribué selon une loi normale de moyenne β et de variance τ^2 . Donc, il est attendu que, pour l'ordonnée par exemple, la valeur de β_0 soit simplement la moyenne des ordonnées prédites β_{0s} (et même chose pour les pentes).

2) La dernière phrase du paragraphe précédent pourrait laisser croire que la régression linéaire devrait donner les mêmes résultats que le modèle MM. Après tout, si le coefficient final est la moyenne des coefficients par sujet, les deux approches devraient donner des résultats similaires, non ? Or, en comparant les Figures 1 et 6, on peut constater que la pente de la régression linéaire ($\beta_1 = 0.363$, $ES = 0.428$, $p = .41$) est environ 3 fois plus petite que la pente moyenne estimée par le MM ($\beta_1 = 1.000$, $ES = 0.213$, $p = .049$). On se souviendra que ce constat est attendu selon la discussion sur la Figure 4, étant explicable par l'estimation *conditionnelle* des effets fixes. Mais il y a un second point qui contribue à la plus grande puissance statistique de l'approche MM : les erreurs standards (l'incertitude autour de cette estimation) sont généralement petites pour les coefficients du MM. Si vous regardez attentivement la Figure 4, vous verrez que les points relatifs à chacune des trajectoires individuelles sont très proches de la droite ajustée (bon ajustement = peu d'incertitude), ce qui contraste avec la droite marginale de la Figure 2 (les points sont très dispersés autour de la droite de régression). Or, le test de signification de chaque coefficient

est simplement $t = \beta/ES_B$. Donc, une erreur standard plus petite a pour conséquence une statistique t plus grande et donc plus significative.

3) Dernier point plus technique : si on compare les trajectoires observées sur les données (en noir) et les trajectoires prédites par le MM (en bleu) à la Figure 4, il est possible de noter qu'elles ne coïncident pas toujours. Pourquoi ? Encore ici, il faut se souvenir que chaque effet aléatoire dans le modèle MM est distribué selon une loi *normale*. Donc, les trajectoires d'origine (ordonnée) et de pente dans la moyenne de la distribution des trajectoires seront bien prédites (il y aura très peu de différences entre la trajectoire observée et prédite) alors que les trajectoires qui s'éloignent beaucoup de la moyenne de la distribution (elles sont donc rares), seront prédites selon la loi normale des effets aléatoires estimée dans nos données. En clair, pour une trajectoire avec une pente beaucoup plus positive que la moyenne des pentes, le modèle MM va prédire une pente un peu moins positive, et vice-versa pour une pente beaucoup moins positive que la moyenne, qui sera prédite comme un peu plus positive. Ce phénomène de "normalisation" des effets aléatoires est attendu et prend le nom de "shrinkage" des effets aléatoires [2, 9].

3.3.4 Effets aléatoires

L'estimation des effets aléatoire est présentée à la Figure 7. La ligne intitulée "Résidu" indique la variance estimée des résidus $Y_{st} - \hat{Y}_{st}$, soit $\hat{\sigma}_\epsilon^2 = 0.856$ (voir Equation 3). Si on compare avec l'estimation de la variance des résidus de la régression linéaire ($\hat{\sigma}^2 = 7.650$), on constate facilement que le modèle multiniveau s'ajuste bien mieux aux données (réduction d'environ 90% de l'erreur du modèle). Clairement, il était avantageux d'inclure l'hétérogénéité des trajectoires dans notre modèle.

Estimations des paramètres de covariance^a

Paramètre		Estimation	Erreur standard	Z de Wald	Sig.	Intervalle de confiance à 95 %	
						Borne inférieure	Borne supérieure
Résidu		,855511	,556343	1,538	,124	,239158	3,060316
Constante + x [sujet = id]	UN (1,1)	5,464393	4,746422	1,151	,250	,995819	29,984955
	UN (2,1)	,703143	,741697	,948	,343	-,750556	2,156843
	UN (2,2)	,090479	,252503	,358	,720	,000381	21,479781

a. Variable dépendante : Variable dépendante.

FIGURE 7 – Modèle MM - effets aléatoires estimés

Les trois autres lignes rapportent les variances et covariances estimées pour les effets aléatoires. On se rappellera que ces effets se distribuent selon une loi normale bivariée avec une certaine forme non-structurée (UN) de matrice de variance-covariance (voir Équation 7). Selon cette équation, UN(1,1) indique la variance à la position 1,1 (rangée 1, colonne 1) de la matrice, soit la variance estimée de l'ordonnée à l'origine : $\hat{\tau}_0^2 = 5.464$. Que signifie ce chiffre intuitivement ? Si vous entrez les $N = 5$ ordonnées prédites selon le MM dans la formule classique de la variance, vous devriez obtenir un estimé assez similaire à cette valeur. Tel que discuté dans la section précédente, on examine plutôt l'intervalle de confiance, qui

indique des valeurs possible (à 95%) entre environ 1 et 30 pour l'ordonnée. N'incluant pas la valeur 0 ($H_0 : \tau_0^2 = 0, H_1 : \tau_0^2 > 0$), on peut rejeter l'hypothèse nulle et conclure que la variance de l'ordonnée est bien *non-nulle*. En clair, il existe bel et bien une diversité significative dans le point de départ (ordonnée à l'origine) des trajectoires estimées.

L'estimé de la variance des pentes est situé à la position 2,2 de la matrice UN : $\hat{\tau}_1^2 = 0.090$. Selon l'intervalle de confiance, il est très probable que la valeur 0 soit incluse (si on regarde la borne inférieure 0.000381), ce qui suggère qu'on ne peut rejeter $H_0 : \tau_1^2 = 0$ avec confiance. On doit donc conclure que les pentes présentent peu ou pas de variabilité entre les sujets, ce qui est cohérent avec la Figure 4.

Enfin, il semble y avoir peu de support sur la présence d'une covariance positive, $UN(2, 1) = \hat{\tau}_{01} = 0.703$ car la valeur 0 est clairement incluse dans l'intervalle de confiance $(-0.750, 2.157)$. Ce résultat suggère que l'ordonnée et la pente des $N = 5$ trajectoires varient de façon indépendante dans cet échantillon.

Quelle est l'implication de tous ces tests sur les effets aléatoires ? Si on veut être parfaitement cohérent, on pourrait estimer à nouveau un modèle MM plus parcimonieux (appelé *random intercept model*, qui est équivalent à un modèle mixte traditionnel), comprenant 4 paramètres au lieu des 6 originaux :

$$Y_{st} = \beta_{0s} + \beta_1 X_{st} + \epsilon_{st} \quad (10)$$

Remarquez que le coefficient relatif à la pente n'a plus d'indice s (ce n'est donc plus un effet aléatoire car il ne varie pas, ce qui implique qu'un seul coefficient sera estimé pour l'ensemble des sujets). Donc, le paramètre relatif à la variance des pentes τ_1^2 sera retiré du modèle. Un second paramètre est également inutile soit la covariance entre l'ordonnée et la pente τ_{01} car (1) ce paramètre est non-significatif et (2) il ne peut y avoir de corrélation entre une variable (ordonnée) et une constante (pente).

En pratique, le choix de retenir ou non certains effets aléatoires sera basé en partie sur leur signification (intervalle de confiance) mais surtout sur les indices d'ajustement du modèle aux données (i.e., les indices AIC ou BIC)[2]. On recherchera le modèle avec le *plus petit* indice d'ajustement, en valeur absolue, pénalisé pour la complexité du modèle (i.e., le nombre de paramètres). De plus, les statisticiens vont souvent distinguer une spécification "structurelle" de l'effet aléatoire (l'effet aléatoire est issu du devis de recherche ou de la collecte de données) vs. "exploratoire" (l'analyste explore si cette variable pourrait être ou non un effet aléatoire mais il n'y a pas d'indication non-équivoque) . Dans le premier cas, la recommandation est de conserver cet effet dans le modèle, indépendamment de sa signification, alors que dans le second cas, les tests de signification et indices devraient guider le choix de retenir ou non ce type d'effet[9].

Attention : il faut se rappeler que les variances ne sont pas comparables (car elles reflètent leur échelle de mesure respective). Donc, il est sans intérêt de commenter, par exemple, " que la variance des ordonnées est plus grande que la variance des pentes" .

4 Modèle multiniveau généralisé

4.1 Introduction aux modèles binaire et ordinal

Certaines situations expérimentales impliquent que les données récoltées pour la variable dépendante ne suivront pas une loi normale. Les observations tirées de quelques lois "non-normales" à la Table 1 sont souvent rencontrées en pratique.

Type de variable Y	Loi statistique (lien $g(Y)$)
Binaire (oui,non)	Binomiale (logit)
Nominale (catégories)	Multinominale (logit généralisé)
Ordinale (catégories ordonnées)	Multinominale (logit cumulatif)
Dénombrement (entier positif)	Poisson / binomiale négative (log)
Continue positive (réel positif)	Exponentielle / gamma (inverse négative)

TABLE 1 – Loi et fonction de lien selon chaque type de variable dépendante

Le modèle de régression multiniveau prendra alors la forme suivante [1, 3], qui est légèrement différente du modèle MM normal (voir Équation 3) :

$$g(Y_{st}) = \beta_{0s} + \beta_{1s}X_{st} \quad (11)$$

où $g(Y)$ est la fonction *généralisée* qui va déterminer la transformation de la variable dépendante originale, afin de permettre l'estimation de la régression multiniveau⁴.

Si on compare avec le modèle MM normal, on peut faire deux constats : (1) les mêmes postulats que ceux du MM normal sont posés pour les effets aléatoires de ce modèle, soit que ceux-ci sont distribués selon une loi normale multivariée (voir Équation 7) ; (2) contrairement à la loi normale qui se définit selon deux paramètres (moyenne et variance), la variance des lois non-normales utilisées ici est fonction de leur moyenne. Donc, la variance des erreurs σ^2 n'étant pas estimée dans un MM généralisé, le terme ϵ_{st} est absent de l'équation 11.

Deux modèles en particulier vont nous intéresser : le modèle multiniveau avec une variable dépendante binaire et celui avec une variable dépendante ordinale. Dans le premier cas, le modèle binaire emprunte directement la définition de la fonction $g(Y)$ de la régression logistique. En fait, on pourrait dire que, d'un point de vue conceptuel, un MM binaire est simplement une régression logistique avec un (ou plusieurs) effet aléatoire. Le modèle prendra donc la forme suivante :

$$g(Y_{st}) = \beta_{0s} + \beta_{1s}X_{st}$$

$$\log \left(\frac{Pr(Y_{st} = 1)}{Pr(Y_{st} = 0)} \right) = \beta_{0s} + \beta_{1s}X_{st} \quad (12)$$

4. Techniquement, un modèle généralisé prend plutôt la forme $g(\mathbb{E}(Y)) = \beta_0 + \beta_1 X + \epsilon$, où $\mathbb{E}(Y)$ est l'*espérance* de la variable dépendante (sa moyenne espérée si on avait un échantillon infini). Toutefois, cette précision mathématique n'est pas importante pour la compréhension de l'usage de cette technique.

où la partie gauche de l'équation 12 est la définition du *logit*, c.a.d., le logarithme du *odds*, celui-ci étant le ratio de la probabilité que $Y = 1$ et de la probabilité que $Y = 0$.

Le deuxième cas est un peu plus complexe. Assumons que nous avons une variable dépendante ordinale qui comprend C catégories ordonnées (p.ex., 1 = 'pas du tout', 2 = 'un peu', ..., C = 'totalement'). Nous ne pouvons pas assumer que la distance entre la valeur 1 et 2 est la même que la distance, par exemple, entre la valeur $C - 1$ et C . Le modèle normal n'est donc pas approprié, même si dans les faits, on utilise souvent celui-ci pour analyser une variable dépendante ordinale (par souci de simplicité ou manque de connaissance sur les modèles généralisés).

Le modèle ordinal utilise une fonction de type "logit cumulatif". Elle est donc de la même forme que la fonction logistique de l'équation 12, mais elle doit tenir compte qu'il y a maintenant C catégories et non seulement 2 (comme dans le modèle binaire). Comment faire? La solution est assez simple : estimer un modèle qui inclut simultanément tous les "changements" de catégories, plutôt qu'un seul comme pour le modèle binaire [1].

Le modèle ordinal prendra donc la forme suivante :

$$\begin{aligned} g(Y_{st}) &= \beta_{0s} + \beta_{1s}X_{st} \\ g(Y_{st}) &= (\beta_0 + u_{0s}) + \beta_{1s}X_{st} \text{ (selon l'équation 5)} \\ \log \left(\frac{Pr(Y_{st} \leq j)}{Pr(Y_{st} > j)} \right) &= \beta_{0j} + u_{0s} + \beta_{1s}X_{st} \end{aligned} \quad (13)$$

pour $j = 1, 2, \dots, C - 1$.

Ce modèle présente deux différences par rapport à l'équation (12). Premièrement, dans la parenthèse du terme gauche, on tente maintenant de prédire la probabilité d'avoir une réponse Y égale ou inférieure à une certaine valeur j , versus la probabilité d'avoir une réponse supérieure. C'est comme si on réalisait une régression logistique où $Y = 1$ si ($Y \leq j$) versus $Y = 0$ si ($Y > j$). Donc, on aura l'équivalent de $C - 1$ régressions logistiques car $j = 1, 2, \dots, C - 1$. Deuxièmement, l'ordonnée à l'origine β_{0j} a maintenant un indice j dans l'équation (13), ce qui suggère que ce paramètre va changer de valeur pour chacune des $C - 1$ régressions logistiques, alors que la partie classique $u_{0s} + \beta_{1s}X_{st}$ du MM restera inchangée. Quelle est l'utilité de permettre à la partie fixe de l'ordonnée de varier selon la j^e régression? Il est nécessaire pour capturer le changement d'ordonnée de la régression en raison du changement de la catégorie de référence (la définition de $Y = 0$ si ($Y > j$) présentée plus haut). En conclusion, les effets fixes estimés dans un tel modèle vont capturer l'effet de X sur le risque de répondre une modalité inférieure de la variable dépendante Y ⁵.

D'un point de vue pratique, il y a peu de difficultés à utiliser les MM généralisés, mis à part le fait qu'on doit être attentif dans un modèle binaire ou ordinal à interpréter le rapport de cote de chaque paramètre (*odds ratio* en anglais, qui se définit comme $\exp(\beta)$), plutôt

5. Si l'intérêt de l'analyste est de prédire le risque de passer à la catégorie *supérieure*, tous les logiciels offrent la possibilité d'inverser l'ordre de modélisation de la variable dépendante, i.e., modéliser de $C, C - 1, \dots, 2, 1$ plutôt que l'ordre initial

que le paramètre β lui-même dans un modèle normal. Le vrai défi est du côté de l'estimation de ces modèles, soit l'étape mathématique et informatique réalisée par le logiciel afin de trouver les valeurs optimales des coefficients de régression.

Les paramètres des MM linéaires (variable dépendante normale) sont relativement faciles à estimer car le modèle au complet (les erreurs de Y et les effets aléatoires) se distribue selon une loi normale multivariée. Donc, les approches classiques de maximum de vraisemblance (comme la méthode REML) convergent vers une solution rapide et optimale. Dans le cas des modèles généralisés, la régression MM inclut une loi non-normale pour la variable dépendante et une loi normale multivariée pour les effets aléatoires. Pour obtenir les paramètres optimaux d'un tel mélange non-linéaire, on doit utiliser une approche d'intégration numérique. La méthode la plus robuste (et la plus populaire), appelée *Adaptive Gaussian Quadrature*, est disponible dans les logiciels SAS et R, mais non dans le logiciel SPSS, qui offre un support très limité pour l'estimation des MM généralisés.

4.2 Exemple SAS pour modèle ordinal

La syntaxe SAS [3] pour estimer ce modèle est présentée dans le tableau suivant. Pour les besoins de la démonstration, la variable dépendante normale Y a été transformée en trois catégories $Y_c \in \{0, 1, 2\}$ selon le tiers de Y .

```
Syntaxe SAS pour MM ordinal

DATA tot;
INPUT id x y yc;
DATALINES;
1 0 10 1
1 1 12 2
1 4 14 2
...
5 3 11 1
5 5 14 2
RUN;

PROC GLIMMIX = tot METHOD = QUAD;
CLASS id;
MODEL yc (DESC) = x / DIST = MULTI LINK = CLOGIT SOLUTION OR(LABEL);
RANDOM intercept x / SUBJECT = id TYPE = UN(1);
COVTEST 'ind' INDEP / PARMS CL;
RUN;
```

La procédure GLIMMIX appelle le modèle mixte généralisé et permet de spécifier la méthode d'estimation numérique `METHOD = QUAD`. L'énoncé `CLASS` permet de spécifier que l'identifiant unique `id` est une variable nominale.

L'énoncé `MODEL` précise la variable dépendante, sa loi, les effets fixes (l'ordonnée n'est pas

précisée car incluse par défaut, comme pour SPSS), et quelques options d’affichage. Le mot-clé DESC après la spécification de la variable dépendante yc indique que l’on désire modéliser la probabilité d’une catégorie supérieure par rapport à une inférieure (le comportement par défaut est inversé, voir équation 13). Le mot-clé DIST = MULTI indique que la variable dépendante se distribue selon une loi multinominale, alors que le mot-clé LINK = CLOGIT spécifie l’usage de la fonction de logit cumulatif, nécessaire pour modéliser une variable dépendante *ordinaire*. Enfin, l’option SOLUTION affiche les valeurs estimés des paramètres fixes alors que l’option OR(LABEL) calcule automatiquement le rapport de cote (Odds Ratio) associé à chaque β .

L’énoncé RANDOM précise les deux effets aléatoires (intercept = β_{0s} et $x = \beta_{1s}$), le mot-clé SUBJECT = id spécifie que la variable pour identifier chaque unité de niveau 2 et TYPE = UN(1) précise la forme de la matrice de variance-covariance des effets aléatoires (voir Équation 7).

Enfin, la dernière ligne débutant avec la commande COVTEST permet d’obtenir les intervalles de confiance à 95% pour chaque variance des effets aléatoires.

La Figure 8 illustre la première partie de la sortie SAS, où sont présentées les informations relatives à la spécification du modèle ordinal.

```

The SAS System
The GLIMMIX Procedure
Model Information
Data Set WORK.TOT
Response Variable yc
Response Distribution Multinomial (ordered)
Link Function Cumulative Logit
Variance Function Default
Variance Matrix Blocked By id
Estimation Technique Maximum Likelihood
Likelihood Approximation Gauss-Hermite Quadrature
Degrees of Freedom Method Containment

Class Level Information
Class Levels Values
id 5 1 2 3 4 5

Number of Observations Read 15
Number of Observations Used 15

Response Profile
Ordered Value yc Total Frequency
1 2 6
2 1 4
3 0 5

The GLIMMIX procedure is modeling the probabilities of levels of
yc having lower Ordered Values in the Response Profile table.

```

FIGURE 8 – Spécification du modèle ordinal

On peut constater, dans la partie intitulée "Model Information", que la variable dépendante est bien spécifiée comme suivant une loi multinominale ordonnée. L’autre point intéressant est l’impact de l’option DESC. Selon le commentaire dans la sortie ("The GLIMMIX procedure is modeling.."), la procedure prédit la "Ordered value = 1 vs 2,3" pour la première

régression, etc. Or, la "Ordered value =1" implique $Y_c = 2$. Donc, cette option permet de modéliser Y_c en ordre inverse : $Pr(Y_c = 2)$ vs. $Pr(Y_c = 0, 1)$ pour la première régression, etc. En clair, chaque β positif pourra s'interpréter comme le risque d'avoir une valeur de Y_c dans la catégorie *supérieure*.

La Figure 9 présente les résultats du modèle multiniveau ordinal. La section intitulée "Covariance Parameters" donne les estimations des variances et covariance des effets aléatoires, alors que la seconde "Solutions for Fixed Effects" présente évidemment les estimations des coefficients de régression des effets fixes.

Covariance Parameter Estimates					
Cov Parm	Subject	Estimate	Standard Error	Wald	95% Confidence Bounds
UN(1,1)	id	18.2786	28.5358	3.3247	80690
UN(2,1)	id	0			
UN(2,2)	id	0.08032	0.8677	0.02267	1.03E184

Solutions for Fixed Effects						
Effect	yc	Estimate	Standard Error	DF	t Value	Pr > t
Intercept	2	-5.1711	4.1551	4	-1.24	0.2813
Intercept	1	-2.4357	3.2463	4	-0.75	0.4948
x		1.5017	1.1339	4	1.32	0.2560

Odds Ratio Estimates				
Comparison	Estimate	DF	95% Confidence Limits	
unit change of x from mean	4.489	4	0.193	104.575

FIGURE 9 – Spécification du modèle ordinal

Pour les variances des effets aléatoires, notons d'abord l'*absence* du terme Residual (si on compare cette sortie avec celle de la Figure 7). Tel qu'expliqué plus haut, il n'y a pas de variance des erreurs σ_e^2 de niveau 1 dans un modèle généralisé. Ensuite, même si les valeurs sont distinctes (rappelons qu'on modélise les 5 trajectoires de la même variable dépendante, mais en version catégorielle à 3 choix de réponses), les constats sont relativement similaires. Premièrement, la variance en position 1,1, soit la variance des ordonnées aléatoires $\tau_0^2 = 18.279$, et sa borne inférieure de l'intervalle de confiance est à 3.3 (donc le zéro en est clairement exclus). Deuxièmement, la variance en position 2,2, soit la variance des pentes aléatoires $\tau_1^2 = 0.080$ présente une borne inférieure très près de zéro. Il semblerait donc encore ici que seules les ordonnées (donc le point de départ des trajectoires) varient significativement entre les 5 sujets.

La partie des effets fixes présente une structure similaire à celle du modèle normal, mais son interprétation est assez différente. Premièrement, on peut constater la présence de 2 ordonnées. Tel qu'expliqué à l'équation (13), il y a une ordonnée pour chacune des $C - 1 = 2$ régressions logistiques à faire pour estimer le modèle ordinal. La première ordonnée, pour $Y_c = 2$, calcule la probabilité de répondre $Y_c = 2$ versus celle de répondre $Y_c < 2$, alors que la seconde ordonnée calcule la $Pr(Y_c = 1, 2)$ versus la $Pr(Y_c = 0)$. Comme 0 est la dernière catégorie, il n'y a pas lieu de faire une 3e régression. Bien qu'il est intéressant de

comprendre la présence de ces estimations d'ordonnées, on interprète rarement ces résultats en pratique. L'intérêt est plutôt pour les prédicteurs X de la probabilité de répondre une catégorie supérieure sur la variable dépendante ordinale.

Le coefficient de régression du prédicteur X (le temps en années) du modèle logistique ordinal est de $\beta_1 = 1.502$. Comme toute régression logistique, ce coefficient s'interprète comme la variation du *logit* pour chaque augmentation d'une unité sur le prédicteur. En clair, ce résultat révèle qu'à chaque année supplémentaire, le logit augmente de 1.502. Est-ce utile? Non car il faut se rappeler que le logit se calcule comme le logarithme du risque. Donc, on interprètera plutôt l'exponentiel du coefficient, $\exp(\beta_1) = 4.489$, qui correspond au fameux rapport de cote (odds ratio). Ici, on peut interpréter plus facilement ce résultat comme : à chaque année supplémentaire (variation d'une unité de X), le risque de rapporter un point de plus à la variable dépendante Y_c (c.a.d., de passer à la catégorie supérieure de la variable ordinale) augmente de 4.49 fois. Est-ce que cette augmentation est statistiquement significative? Si on examine l'intervalle de confiance présenté dans la section "Odds Ratio Estimates", la réponse est non car les valeurs probables (à 95%) du rapport de cote varient entre 0.193 et 104.575, ce qui inclut la valeur 1, soit le rapport de cote nul (un risque de 1 = pas de changement du risque).

Références

- [1] Agresti, A. (2002). *Categorical Data Analysis* (2nd ed.). New-York : Wiley. [chapitre 12 sur les modèles mixtes pour variable dépendante binaire ou ordinale]
- [2] Brown, H., & Prescott, R. (2006). *Applied mixed models in medicine* (2nd ed.). New-York : J. Wiley & Sons. [superbe introduction peu mathématique aux modèles multiniveaux linéaires]
- [3] Molenberghs, G., & Verbeke, G. (2005). *Models for discrete longitudinal data*. NYC : Springer. [chapitres 14-16 sur les modèles mixtes pour variable dépendante binaire ou ordinale]
- [4] Peugh, J. L., & Enders, C. K. (2005). Using the SPSS Mixed procedure to fit cross-sectional and longitudinal multilevel models. *Educational and Psychological Measurement*, 65(5), 717-741. [article non-mathématique qui refait les analyses multiniveaux de Singer & Willett (2003) avec SPSS]
- [5] Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical Linear Models : Applications and Data Analysis Methods* (2nd ed.). Thousand Oaks, CA : Sage Publications. [livre assez mathématique qui introduit la notation multiniveaux - un classique sur le sujet]
- [6] Singer, J. D., & Willett, J. B. (2003). *Applied Longitudinal Data Analysis : Modeling Change and Event Occurrence*. NYC : Oxford University Press. [les chapitres 1 à 6 sont une superbe introduction non-mathématique aux modèles multiniveaux longitudinaux]
- [7] Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel Analysis : An Introduction to Basic and Advanced Multilevel Modeling* (2nd ed.). Washington, DC : Sage Publications. [livre avancé sur les aspects plus pratiques des modèles multiniveaux]
- [8] SPSS. (2005). *Linear Mixed-Effects Modeling in SPSS : An Introduction to the MIXED Procedure* : SPSS Corporation. [introduction accessible à la procédure SPSS MIXED]
- [9] Verbeke, G., & Molenberghs, G. (2009). *Linear Mixed Models for Longitudinal Data*. NYC : Springer. [la bible statistique sur les modèles mixtes linéaires]