

2m11.2968.1

Université de Montréal

Analyse longitudinale à l'aide de la méthode des  
équations d'estimation généralisées pour de petits  
échantillons.

par

**Rolland Gaudet**

Département de mathématiques et de statistique  
Faculté des arts et des sciences

Mémoire présenté à la Faculté des études supérieures  
en vue de l'obtention du grade de  
Maître ès sciences (M.Sc.)  
en statistique

février 2002

© Rolland Gaudet, 2002



QA

3

U54

2002

V.012

# Université de Montréal

Faculté des études supérieures

Ce mémoire intitulé

## Analyse longitudinale à l'aide de la méthode des équations d'estimation généralisées pour de petits échantillons.

présenté par

**Rolland Gaudet**

a été évalué par un jury composé des personnes suivantes :

*Robert Cléroux*

---

(président-rapporteur)

*Yves Lepage*

---

(directeur de recherche)

*Narayan C. Giri*

---

(membre du jury)

Mémoire accepté le:

---

## SUMMARY

---

This mémoire is an analysis of a regression method for longitudinal data: The Generalized Estimating Equations method (GEE).

Primarily, we will have a theoretical overlook of a few regression methods to then know how to do regression with longitudinal data.

After having explained in theory the GEE method, we will display a review of the literature regarding this method used with small samples.

Finally, we will do the entire analysis of this method, in terms of statistics, with the use of Monte Carlo type simulations.

# SOMMAIRE

---

Ce mémoire se veut une analyse d'une méthode de régression pour données longitudinales: la méthode des équations d'estimation généralisées (EEG).

Premièrement, nous ferons un survol théorique de quelques méthodes de régression pour en arriver à voir comment faire de la régression avec des données longitudinales.

Après avoir expliqué de manière théorique la méthode des EEG, nous ferons une revue de la littérature de cette méthode utilisée avec de petits échantillons.

Finalement, nous ferons une analyse complète de cette méthode au niveau statistique à l'aide de simulation de type Monte Carlo.

# REMERCIEMENTS

---

Je tiens à remercier sincèrement mon directeur de recherche M. Yves Lepage pour sa compétence et sa disponibilité. Je remercie aussi M. Jean Lambert, avec qui j'ai débuté ce mémoire.

Je dédis ce mémoire à ma femme, Julie, et mes filles, Béatrice et Laurence, qui ont su m'endurer et m'appuyer durant ces longues années.

Ma mère et mon père m'ont aussi offert un soutien moral et financier très appréciable.

Sans vous tous, la rédaction de ce mémoire aurait été passablement plus difficile.

Encore une fois MERCI.

# Table des matières

---

<b>Sommaire</b> .....	iii
<b>Remerciements</b> .....	iv
<b>Liste des tableaux</b> .....	vii
<b>Introduction</b> .....	1
<b>Chapitre 1. Contexte théorique</b> .....	2
1.1. Régression par moindres carrés.....	2
1.2. Modèle linéaire généralisé .....	7
1.3. Analyse longitudinale .....	9
1.3.1. Notation .....	11
1.3.2. Contexte de l'étude longitudinale .....	12
1.4. Méthode des équations d'estimation généralisées (EEG) .....	13
1.4.1. Quasi-vraisemblance .....	13
1.4.2. Développement théorique des EEG.....	15
1.4.3. Structure de corrélation .....	18
1.4.4. Efficacité des EEG.....	21
1.5. Revue de la littérature .....	23
1.5.1. Conclusion.....	32
<b>Chapitre 2. Analyse des résultats</b> .....	34
2.1. La simulation .....	34

2.2.	Analyse des résultats.....	38
2.3.	Conclusion.....	47
	<b>Annexe A. Données binaires corrélées.....</b>	<b>49</b>
	<b>Annexe B. Tableaux de résultats.....</b>	<b>51</b>
	<b>Bibliographie.....</b>	<b>77</b>



## Liste des tableaux

---

Tableau 1.1.....	8
Tableau 1.2.....	22
Tableau 1.3.....	24
Tableau 1.4.....	26
Tableau 1.5.....	27
Tableau 1.6.....	29
Tableau 1.7.....	30
Tableau 1.8.....	31
Tableau 2.1.....	35
Tableau 2.2.....	43
Tableau 2.3.....	46
Tableau 3.1.....	53
Tableau 3.2.....	54
Tableau 3.3.....	55
Tableau 3.4.....	56
Tableau 4.1.....	57
Tableau 4.2.....	58
Tableau 4.3.....	59
Tableau 4.4.....	60
Tableau 5.1.....	61
Tableau 5.2.....	62
Tableau 5.3.....	63
Tableau 5.4.....	64
Tableau 6.1.....	65
Tableau 6.2.....	66
Tableau 6.3.....	67
Tableau 6.4.....	68
Tableau 7.1.....	69
Tableau 7.2.....	70
Tableau 7.3.....	71
Tableau 7.4.....	72
Tableau 8.1.....	73
Tableau 8.2.....	74
Tableau 8.3.....	75
Tableau 8.4.....	76

# INTRODUCTION

---

Dans le domaine de la santé, nous sommes souvent confrontés à de petits échantillons. L'analyse longitudinale est d'autant plus exigeante puisqu'il nous faut des observations à plusieurs instants dans le temps pour chaque sujet.

Il est démontré (Liang & Zeger (1986)) que la méthode des équations d'estimation généralisées (EEG) donne des estimateurs sans biais suivant asymptotiquement une loi multinormale mais qu'en est-il pour de petits échantillons?

Nous allons donc analyser les estimateurs de la régression de la méthode des EEG en fonction du biais relatif, de la variance et de l'étendue. Pour ce faire nous avons simulé des données corrélées avec une certaine structure de corrélation en faisant varier le nombre de sujets, le nombre de répétitions par sujet, le niveau de corrélation, l'estimateur de la variance et ce pour des données binaires et de loi normale.

# Chapitre 1

---

## CONTEXTE THÉORIQUE

### 1.1. RÉGRESSION PAR MOINDRES CARRÉS

La régression linéaire est possiblement la technique statistique la plus utilisée pour établir des relations linéaires entre deux ou plusieurs variables.

Pour un ensemble de  $s$  observations, la relation entre une variable dépendante et  $p-1$  ( $p > 1$ ) variables explicatives ou indépendantes est représentée par le modèle de régression linéaire suivant

$$y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_{p-1} x_{ip-1} + \epsilon_i, \quad p > 1, \quad i = 1, \dots, s,$$

où  $y_i$  représente la valeur de la variable dépendante du  $i^{\text{ième}}$  sujet,  $x_{ij}$  représente la valeur de la  $j^{\text{ième}}$  variable indépendante pour le  $i^{\text{ième}}$  sujet,  $\epsilon_i$  représente l'erreur aléatoire associée au  $i^{\text{ième}}$  sujet et les  $\beta_k$  ( $k=0, \dots, p-1$ ) représentent les coefficients de régression à estimer. Nous supposons que les erreurs aléatoires,  $\epsilon_i$ , sont indépendantes et identiquement distribuées selon la loi normale de moyenne 0 et de variance constante ( $\sigma^2(\epsilon_i) = \sigma^2$ ), c'est-à-dire

$$\epsilon_i \sim N(0, \sigma^2), i = 1, \dots, s.$$

Sous forme matricielle, nous avons

$$\underline{Y} = X\underline{\beta} + \underline{\epsilon}$$

où

$$\begin{aligned}\underline{Y} &= (y_1, y_2, \dots, y_s)'_{s \times 1}, \\ X &= \begin{bmatrix} 1 & x_{11} & \dots & x_{1p-1} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 1 & x_{s1} & \dots & x_{sp-1} \end{bmatrix} = \begin{bmatrix} \underline{X}_1 \\ \cdot \\ \cdot \\ \cdot \\ \underline{X}_s \end{bmatrix}_{s \times p}, \\ \underline{\beta} &= (\beta_0, \beta_1, \dots, \beta_{p-1})'_{p \times 1}, \\ \underline{\epsilon} &= (\epsilon_1, \epsilon_2, \dots, \epsilon_s)'_{s \times 1}\end{aligned}$$

avec  $\underline{\epsilon}$  de loi multinormale  $N_s(\underline{0}, \sigma^2 I_s)$ ,  $\underline{0}$  étant le vecteur nul de dimension  $s \times 1$  et  $I_s$  la matrice identité  $s \times s$ .

Le modèle s'écrit alors

$$E(y_i) = \beta_0 + \beta_1 x_{i1} + \dots + \beta_{p-1} x_{ip-1} = \underline{X}_i \underline{\beta}, \quad i = 1, \dots, s.$$

L'estimateur selon la méthode des moindres carrés,  $\hat{\underline{\beta}}$ , est obtenu en solutionnant l'équation normale

$$X'(\underline{Y} - X\underline{\beta}) = 0$$

et en supposant  $X$  de plein rang, l'estimateur

$$\hat{\underline{\beta}} = (X'X)^{-1}X'\underline{Y} \tag{1.1.1}$$

est unique.

Le vecteur des valeurs prédites est  $\hat{\underline{Y}} = X\hat{\underline{\beta}}$  et le vecteur des résidus ( $\underline{e}$ ) correspond à

$$\underline{e} = \underline{Y} - \hat{\underline{Y}} = \underline{Y} - X\hat{\underline{\beta}}.$$

Ainsi, l'estimateur sans biais de  $\sigma^2$ , noté  $S^2$ , est défini par

$$S^2 = \frac{\underline{e}'\underline{e}}{(s-p)} = \frac{(\underline{Y} - X\underline{\hat{\beta}})'(\underline{Y} - X\underline{\hat{\beta}})}{(s-p)}.$$

L'estimateur  $\underline{\hat{\beta}}$  de  $\underline{\beta}$  est sans biais pour  $\underline{\beta}$ , c'est-à-dire  $E(\underline{\hat{\beta}}) = \underline{\beta}$ . De plus, parmi tous les estimateurs sans biais de  $\underline{\beta}$ ,  $\underline{\hat{\beta}}$  possède la plus petite variance, c'est-à-dire que  $\underline{\hat{\beta}}$  est un estimateur sans biais à variance minimale. La matrice de covariances de  $\underline{\hat{\beta}}$  est donnée par

$$\text{Var}(\underline{\hat{\beta}}) = \sigma^2(X'X)^{-1} \quad (1.1.2)$$

et l'estimateur de cette matrice de covariances est obtenu en utilisant  $S^2$

$$\widehat{\text{Var}}(\underline{\hat{\beta}}) = S^2(X'X)^{-1}.$$

L'estimateur  $\underline{\hat{\beta}}$  suit une loi multinormale de moyenne  $\underline{\beta}$  et de matrice de covariances définie à l'équation (1.1.2).

Remarquons que l'estimateur des moindres carrés est identique à l'estimateur du maximum de vraisemblance lorsque les variables aléatoires  $\epsilon_i$  sont indépendantes et suivent une loi normale. Pour plus de détails sur la régression linéaire, nous pouvons consulter Neter, Wasserman & Kutner (1990).

Une forte dépendance entre la variable dépendante et un groupe de variables indépendantes assure une relation entre ces deux groupes. D'autre part, la variance des  $\epsilon_i$  ( $i=1, \dots, s$ ) doit être constante (homocédasticité) pour que les estimateurs des coefficients de régression possèdent une variance non biaisée. Lorsque la variance de l'erreur aléatoire n'est pas constante, une transformation de la variable dépendante est envisageable mais nous pourrions alors modifier la relation

que nous tentons d'expliquer entre les deux groupes de variables.

Une alternative consiste à utiliser la régression des moindres carrés pondérés. Cette technique minimise une somme pondérée du carré des résidus ( $e_i$ ) construite en associant un poids  $\omega_i$  à la  $i^{\text{ème}}$  valeur de la variable dépendante ( $y_i$ ) et des variables indépendantes ( $x_{i1}, x_{i2}, \dots, x_{ip-1}$ ).

La matrice des poids,  $W$ , est donnée par

$$W = \begin{bmatrix} \omega_1 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & \omega_2 & 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \omega_{s-1} & 0 \\ 0 & \cdot & \cdot & \cdot & 0 & \omega_s \end{bmatrix}_{s \times s}$$

où les  $\omega_i$  sont des constantes positives définies par  $\text{Var}(\epsilon_i) = \frac{\sigma^2}{\omega_i}$ ,  $i=1, \dots, s$ ,  $\sigma^2$  étant une constante positive. Par la suite, nous transformons la variable dépendante et les variables indépendantes comme suit

$$y_i^* = y_i \sqrt{\omega_i},$$

$$x_{ij}^* = x_{ij} \sqrt{\omega_i}.$$

Maintenant toutes les valeurs de la variable  $\underline{Y}^*$  ont la même variance ( $\sigma^2$ ) dépendamment de la façon dont  $\omega_i$  est défini. Il suffit donc d'appliquer la méthode des moindres carrés aux variables  $\underline{Y}^*$  et  $X^*$ . La somme du carré des écarts entre les valeurs observées  $y_i^*$  et de leur moyenne  $E(y_i^*)$  est donnée, sous forme matricielle, par

$$(\underline{Y}^* - X^* \underline{\beta}^*)' (\underline{Y}^* - X^* \underline{\beta}^*).$$

L'estimateur  $\underline{\hat{\beta}}^*$  est la solution de l'équation normale suivante

$$X'(W\underline{Y} - WX\underline{\beta}^*) = 0$$

c'est-à-dire,

$$\underline{\hat{\beta}}^* = (X'WX)^{-1}X'W\underline{Y}. \quad (1.1.3)$$

Notons que si  $\omega_i = \omega_j$  pour tout  $i \neq j$ ,  $i$  et  $j=1, \dots, s$ , alors l'équation (1.1.3) se réduit à l'équation (1.1.1). La matrice de covariances de cet estimateur est

$$\text{var}(\underline{\hat{\beta}}^*) = \sigma^2(X'WX)^{-1} \quad (1.1.4)$$

et son estimateur est

$$\widehat{\text{var}}(\underline{\hat{\beta}}^*) = S^{*2}(X'WX)^{-1}$$

où

$$S^{*2} = \frac{\sum_{i=1}^s w_i (y_i - \hat{y}_i)^2}{s - p}$$

et

$$\hat{Y} = (\hat{y}_1, \dots, \hat{y}_p).$$

L'estimateur  $\underline{\hat{\beta}}^*$  suit une loi multinormale de moyenne  $\underline{\beta}^*$  et de matrice de covariances définie à l'équation (1.1.4).

La méthode des moindres carrés pondérés a été développée il y a de nombreuses années et nous pouvons trouver plus de détails dans Neter, Wasserman & Kutner (1990).

## 1.2. MODÈLE LINÉAIRE GÉNÉRALISÉ

Ce type de modèle utilise une fonction de liaison (G) monotone et différentiable pour relier  $E(y_i) = \mu_i$  aux variables indépendantes  $\underline{X}_i\beta$ , c'est-à-dire que l'on considère le modèle

$$G(\mu_i) = \underline{X}_i\beta, \quad i = 1, \dots, s.$$

Les deux modèles linéaires de la section précédente utilisaient  $\mu_i = X_i\beta$ . Cette méthode exige aussi que la variance des  $y_i$  soit une fonction connue (M) de son espérance, c'est-à-dire

$$Var(y_i) = \sigma^2 M(\mu_i) \quad i = 1, \dots, s,$$

où  $\sigma^2$  est une constante.

Dans le tableau 1.1, nous retrouvons une liste non exhaustive des fonctions de liaison et des fonctions de la variance pour des distributions appartenant à la famille des distributions exponentielles. En pratique, il nous faut donc choisir les deux fonctions (G et M) selon la distribution de la variable dépendante.



TABLEAU 1.1

Fonction de liaison et fonction de la variance  
pour le modèle linéaire généralisé

Distribution	Fonction de liaison (G)	Fonction de la variance (M)
Normale	$G(\mu_i) = \mu_i = \underline{X}_i \underline{\beta}$	$M(\mu_i) = 1$
Poisson	$G(\mu_i) = \log(\mu_i) = \underline{X}_i \underline{\beta}$	$M(\mu_i) = \mu_i$
Binômiale	$G(\mu_i) = \log(\mu_i/(1 - \mu_i)) = \underline{X}_i \underline{\beta}$	$M(\mu_i) = \mu_i(1 - \mu_i)$
Gamma	$G(\mu_i) = \mu_i^{-1} = \underline{X}_i \underline{\beta}$	$M(\mu_i) = \mu_i^2$

Toujours concernant la variable dépendante, la matrice de covariances de  $\underline{Y}$  est de la forme

$$\sigma^2(\underline{Y}) = \sigma^2 M = \sigma^2 \begin{bmatrix} M(\mu_1) & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & M(\mu_2) & 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & 0 & \cdot \\ 0 & \cdot & \cdot & \cdot & 0 & M(\mu_s) \end{bmatrix}_{s \times s}$$

Étant donné des observations indépendantes ou du moins non-corrélées (McCullagh & Nelder (1989)), l'estimateur  $\hat{\underline{\beta}}$  de  $\underline{\beta}$ , selon la méthode des modèles linéaires généralisés, est la solution de l'équation

$$D' M^{-1}(\underline{Y} - X \underline{\beta}) = \underline{0}$$

où  $X\underline{\beta} = G(\underline{\mu})$ ,  $\underline{\mu} = (\mu_1, \dots, \mu_s)'$  et

$$D = \frac{\partial \underline{\mu}}{\partial \underline{\beta}} = \begin{bmatrix} \frac{\partial \mu_1}{\partial \beta_0} & \dots & \frac{\partial \mu_1}{\partial \beta_{p-1}} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \frac{\partial \mu_s}{\partial \beta_0} & \dots & \frac{\partial \mu_s}{\partial \beta_{p-1}} \end{bmatrix}_{s \times p}$$

L'estimateur  $\hat{\underline{\beta}}$  suit asymptotiquement une loi multinormale de moyenne  $\underline{\beta}$  et de matrice de covariances

$$\sigma^2 [D' M^{-1} D]^{-1}.$$

Ces résultats sur les modèles linéaires généralisés se retrouvent dans plusieurs articles et livres de statistique dont en particulier Dunlop (1994).

### 1.3. ANALYSE LONGITUDINALE

L'analyse longitudinale s'effectue avec des sujets dont les observations sont généralement répétées dans le temps. Les techniques statistiques énoncées dans les deux sections précédentes ne sont donc pas applicables à des observations provenant d'études longitudinales étant donné la dépendance des observations dans le temps. D'ailleurs, Dunlop (1994) l'a bien illustré en étudiant deux modèles de régression linéaire simple. Le premier modélise l'effet de groupe tandis que le deuxième modélise l'effet du temps.

Considérons un nombre  $n$  pair de sujets et notons par  $y_{it}$  la valeur mesurée pour le  $i^{\text{ième}}$  sujet au temps  $t$  ( $t=0,1$ ). Le modèle de régression est donc

$$y_{it} = \beta_0 + \beta_1 x_i + \epsilon_{it}$$

où  $\epsilon_{it}$  sont des erreurs aléatoires telles que  $\text{corr}(\epsilon_{i0}, \epsilon_{i1}) = \rho$  ( $\rho \neq 0$ ),  $i = 1, \dots, s$  et  $t = 0, 1$ . Supposons que la variance des erreurs est  $\sigma^2$  et que les erreurs inter-sujets sont indépendantes ( $\text{corr}(\epsilon_{it}, \epsilon_{jt}) = 0$  pour  $i \neq j$ ). Supposons aussi que  $x_i = 0$  pour  $i = 1, \dots, \frac{s}{2}$  et  $x_i = 1$  pour  $i = \frac{s}{2} + 1, \dots, s$ . Le coefficient  $\beta_0$  représente alors l'effet moyen pour les  $\frac{s}{2}$  premiers sujets et  $\beta_0 + \beta_1$  représente l'effet moyen des  $\frac{s}{2}$  derniers sujets. Le coefficient  $\beta_1$  représente donc l'effet de groupe.

Selon Dunlop (1994), l'estimateur des moindres carrés de ce coefficient est

$$\hat{\beta}_1 = \frac{\sum_i \sum_t (y_{it} - \bar{y})(x_i - \bar{x})}{\sum_i \sum_t (x_i - \bar{x})^2} = \frac{1}{s} \left[ \sum_{i=\frac{s}{2}+1}^s (y_{i0} + y_{i1}) - \sum_{i=1}^{\frac{s}{2}} (y_{i0} + y_{i1}) \right]$$

et sa variance est

$$\text{Var}(\hat{\beta}_1) = \text{Var}\left(\frac{1}{s} \left[ \sum_{i=\frac{s}{2}+1}^s (y_{i0} + y_{i1}) - \sum_{i=1}^{\frac{s}{2}} (y_{i0} + y_{i1}) \right]\right) = \frac{2\sigma^2}{s} (1 + \rho).$$

Pour une corrélation intra-sujet positive ( $\rho > 0$ ), nous remarquons que si nous ne tenons pas compte de la corrélation intra-sujet, nous sous-estimons la variance de l'estimateur  $\hat{\beta}_1$ . Cela a pour effet de produire des tests statistiques plus libéraux et des intervalles de confiance plus courts. Une interprétation analogue est possible si  $\rho < 0$ .

Modélisons maintenant l'effet du temps avec le modèle suivant

$$y_{it} = \beta_0^* + \beta_1^* t + \epsilon_{it}^* \quad i = 1, \dots, s, \quad t = 0, 1.$$

Le coefficient  $\beta_1^*$  représente donc l'effet du temps. Imposons les mêmes conditions que pour le modèle précédent quant aux corrélations intra-sujet et inter-sujets.

Selon Dunlop (1994), l'estimateur des moindres carrés de ce coefficient est

$$\hat{\beta}_1^* = \frac{\sum_i \sum_t (y_{it} - \bar{y})(x_i - \bar{x})}{\sum_i \sum_t (x_i - \bar{x})^2} = \frac{1}{s} \sum_i (y_{i1} - y_{i0})$$

et sa variance est

$$Var(\hat{\beta}_1^*) = Var\left(\frac{1}{s} \sum_i (y_{i1} - y_{i0})\right) = \frac{2\sigma^2}{s}(1 - \rho).$$

Nous remarquons alors une surestimation de la variance de l'estimateur  $\hat{\beta}_1^*$  si nous ne tenons pas compte de la corrélation intra-sujet positive. Cela nous conduit à des tests conservateurs et des intervalles de confiance plus longs. Une interprétation analogue est possible lorsque  $\rho < 0$ .

D'une manière générale, avec plus de deux temps, la variance d'un coefficient de régression modélisant un effet de groupe est sous-estimée et la variance d'un coefficient modélisant un effet de temps est surestimée si la corrélation intra-sujet positive est ignorée (Zeger (1988)). La régression linéaire n'est donc pas à utiliser.

### 1.3.1. Notation

Établissons  $y_{it}$  comme étant la valeur de la variable dépendante du  $i^{\text{ième}}$  sujet ( $i=1, \dots, s$ ) au temps  $t$  ( $t=1, \dots, n_i$ ) pour un grand total de  $N = \sum_{i=1}^s n_i$  observations et le vecteur de la variable indépendante ainsi que la matrice des  $p-1$  ( $p > 1$ ) variables

indépendantes du  $i^{\text{ième}}$  sujet comme étant

$$\underline{Y}_i = (y_{i1}, \dots, y_{in_i})'_{n_i \times 1}$$

et

$$X_i = \begin{bmatrix} 1 & x_{i11} & \dots & x_{i,p-1,1} \\ 1 & x_{i12} & \dots & x_{i,p-1,2} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 1 & x_{i1n_i} & \dots & x_{i,p-1,n_i} \end{bmatrix}_{n_i \times p}$$

où  $x_{ijt}$  est une observation obtenue du  $i^{\text{ième}}$  sujet ( $i=1, \dots, s$ ) pour la  $j^{\text{ième}}$  variable indépendante ( $j=1, \dots, p-1$ ) au temps  $t$  ( $t=1, \dots, n_i$ ). L'équivalent pour les  $s$  sujets est

$$\underline{Y} = (\underline{Y}_1, \dots, \underline{Y}_s)'_{N \times 1}$$

et

$$X = \begin{bmatrix} X_1, \dots, X_s \end{bmatrix}'_{N \times p}$$

### 1.3.2. Contexte de l'étude longitudinale

Dans le cadre d'une analyse longitudinale, nous avons un vecteur de mesures pour la variable dépendante ainsi que pour la (les) variable(s) indépendante(s) et ce pour chaque sujet. Si plus d'une variable indépendante est utilisée, les vecteurs seront juxtaposés pour former une matrice. Chacun de ces vecteurs est en fait une répétition des  $n_i$  mesures d'une même variable, observées à des temps différents

et, idéalement, uniformément espacées dans le temps.

Tout au long de ce travail nous nous concentrerons sur une généralisation du modèle marginal (Diggle, Liang & Zeger (1994)) pour l'analyse de données provenant d'études longitudinales soit, les équations d'estimation généralisées.

#### 1.4. MÉTHODE DES ÉQUATIONS D'ESTIMATION GÉNÉRALISÉES (EEG)

Cette méthode, développée par Liang & Zeger (1986), tient compte de la dépendance intra-sujet des observations. Les auteurs ont utilisé la quasi-vraisemblance (McCullagh & Nelder (1983)) pour développer une méthode d'estimation itérative particulière afin de modéliser des observations provenant d'études longitudinales.

##### 1.4.1. Quasi-vraisemblance

L'utilisation de cette méthode itérative ne requiert que peu de présupposés par rapport à la distribution de la variable dépendante. Nous devons seulement spécifier, comme pour les modèles linéaires généralisés, la fonction linéaire entre la moyenne de la variable dépendante et les covariables ainsi que la fonction entre la moyenne et la variance de la variable dépendante. Par contre, la variance des estimateurs des coefficients de la régression ne dépend pas des autres paramètres qui sont utilisés pour calculer ces estimateurs. C'est la seule différence qu'il y a entre la méthode itérative de la quasi-vraisemblance et la méthode de la vraisemblance utilisée par les modèles linéaires généralisés.

Considérons le modèle linéaire généralisé défini à la section 1.2

$$G(\underline{\mu}_i) = X_i \underline{\beta}$$

où  $E(\underline{Y}_i) = \underline{\mu}_i$

La matrice de covariances de  $\underline{Y}_i$ ,  $U_i$ , est exprimée à l'aide d'une fonction matricielle connue (M) de la moyenne  $\underline{\mu}_i$  et un paramètre d'échelle  $\phi$ , c'est-à-dire

$$U_i = M(\underline{\mu}_i)/\phi, i = 1, \dots, s. \quad (1.4.1)$$

Le paramètre d'échelle peut être estimé, pour une itération donnée, à partir des résidus standardisés

$$\hat{\gamma}_{it} = \frac{y_{it} - \hat{y}_{it}}{\sqrt{Var(y_i)}} \quad (1.4.2)$$

alors

$$\hat{\phi}^{-1} = \sum_{i=1}^s \sum_{t=1}^{n_i} \frac{\hat{\gamma}_{it}^2}{N - p + 1} \quad (1.4.3)$$

où  $p-1$  est le nombre de variables indépendantes et  $N = \sum_{i=1}^s n_i$ .

L'estimateur  $\hat{\underline{\beta}}$  de la quasi-vraisemblance est la solution du système d'équations suivant

$$\sum_{i=1}^s D_i' U_i^{-1} (\underline{Y}_i - \underline{\mu}_i) = \underline{0}. \quad (1.4.4)$$

où  $D_i = \frac{\partial \underline{\mu}_i}{\partial \underline{\beta}} : n_i * p$ .

L'estimateur ainsi trouvé,  $\hat{\underline{\beta}}$ , suit asymptotiquement, sous certaines conditions de régularité (McCullagh (1983)), une loi multinormale de moyenne  $\underline{\beta}$  et de matrice de covariances

$$\sigma^2 [D' M^{-1} D]^{-1}$$

où

$$D = \begin{bmatrix} D_1 \\ D_2 \\ \cdot \\ \cdot \\ D_{s-1} \\ D_s \end{bmatrix}_{N \times p}$$

#### 1.4.2. Développement théorique des EEG

Soit  $R_i(\rho)$  la matrice de corrélation du  $i^{\text{ième}}$  sujet,  $i=1, \dots, s$ . C'est par cette matrice que le modèle de régression tient compte de la corrélation intra-sujet.

Le modèle utilisé est le même que pour les modèles linéaires généralisés, c'est-à-dire

$$G(\underline{\mu}_i) = X_i \underline{\beta}$$

où  $E(\underline{Y}_i) = \underline{\mu}_i$ ,  $i=1, \dots, s$ . La matrice de covariances des  $\underline{Y}_i$  est de la forme  $\sigma^2 U_i^*$  où

$$U_i^* = M(\underline{\mu}_i)^{1/2} R_i(\rho) M(\underline{\mu}_i)^{1/2} / \phi \quad (1.4.5)$$

et

$$M(\underline{\mu}_i) = \begin{bmatrix} m(\mu_{i1}) & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & m(\mu_{i2}) & 0 & \cdot & \cdot & 0 \\ \cdot & & \cdot & & & \cdot \\ \cdot & & & \cdot & & \cdot \\ \cdot & & & & & 0 \\ 0 & \cdot & \cdot & \cdot & 0 & m(\mu_{in_i}) \end{bmatrix}_{n_i \times n_i},$$



$i=1, \dots, s$ . Nous pouvons estimer  $\phi$  à l'aide de l'équation (1.4.3). Remarquons que l'équation (1.4.5) se réduit à l'équation (1.4.1) si  $n_i = 1$  pour  $i=1, \dots, s$ .

La matrice de covariances pour les  $s$  sujets est une matrice diagonale par blocs due à l'indépendance inter-sujets, c'est-à-dire

$$\sigma^2 U^* = \sigma^2 \begin{bmatrix} U_1^* & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & U_2^* & 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & \cdot & \cdot & \cdot & 0 & U_s^* \end{bmatrix}_{N \times N}$$

A cause de cette particularité de la matrice  $\sigma^2 U^*$ , l'estimateur de  $\underline{\beta}$  selon la méthode des EEG est la solution du système d'équations suivant

$$\sum_{i=1}^s D_i' U_i^{*-1} (\underline{Y}_i - \underline{\mu}_i) = \underline{0} \quad (1.4.6)$$

où  $D_i = \frac{\partial \underline{\mu}_i}{\partial \underline{\beta}} : n_i * p$ .

Nous remarquons que l'équation (1.4.6) se distingue de l'équation (1.4.4) par le fait que, contrairement à  $U_i$  qui est une matrice diagonale,  $U_i^*$  n'est pas une matrice diagonale afin de permettre une corrélation intra-sujet.

Le système d'équations d'estimation (1.4.6) est résolu en utilisant la procédure itérative suivante:

1) A partir de valeurs initiales pour la matrice de corrélation  $R_i(\rho)$  et du paramètre d'échelle  $\phi$ , nous utilisons la méthode des moindres carrés pondérés pour

estimer  $\underline{\beta}$ .

- 2) Avec la valeur de l'estimateur ( $\hat{\beta}$ ) de  $\beta$ , nous calculons les résidus.
- 3) Nous utilisons les résidus pour réestimer  $\hat{\phi}$  et  $R_i(\hat{\rho})$  (voir section 1.4.3) .
- 4) Ces trois étapes sont répétées jusqu'à ce qu'un critère de convergence pour les estimateurs soit atteint.

Sous certaines conditions de régularité, Liang & Zeger (1986) ont démontré que lorsque  $n \rightarrow \infty$ , les estimateurs trouvés à partir de l'équation (1.4.6) sont consistants et  $\sqrt{s}(\hat{\beta} - \underline{\beta})$  suit asymptotiquement une loi multinormale de moyenne  $\underline{0}$  et de matrice de covariances

$$V = \lim_{s \rightarrow \infty} s \frac{[\sum_{i=1}^s D_i' U_i^{*-1} cov(\underline{Y}_i) U_i^{*-1} D_i]}{(\sum_{i=1}^s D_i' U_i^{*-1} D_i) (\sum_{i=1}^s D_i' U_i^{*-1} D_i)} \quad (1.4.7)$$

où  $cov(\underline{Y}_i)$  est la matrice de covariances des  $n_i$  répétitions de la variable dépendante pour le  $i^{ième}$  sujet. Nous pouvons estimer la matrice  $V$  en remplaçant  $cov(\underline{Y}_i)$  par  $\underline{h}_i \underline{h}_i'$ , où  $\underline{h}_i = (\underline{Y}_i - \underline{\mu}_i)$  et en substituant  $\rho$ ,  $\underline{\beta}$  et  $\phi$  par leurs estimateurs respectifs.

Prentice (1988) a généralisé l'estimateur de la variance de  $\underline{\beta}$  de Liang & Zeger (1986) pour une variable indépendante binaire. Il a, sous certaines conditions de régularité, développé la distribution conjointe asymptotique de  $\sqrt{s}(\hat{\beta} - \underline{\beta})$  et  $\sqrt{s}(\hat{\rho} - \underline{\rho})$ . De ce fait, il a pu remplacer l'estimateur de la matrice de covariances de  $Y_i$  ( $\widehat{cov}(\underline{Y}_i)$ ) par  $U_i^*$  pour obtenir un estimateur de la variance de  $\sqrt{s}(\hat{\beta} - \underline{\beta})$  beaucoup plus simple que celui de Liang & Zeger (1986), c'est-à-dire

$$s \left( \sum_{i=1}^s D_i' U_i^{*-1} D_i \right)^{-1}. \quad (1.4.8)$$

Emrich & coll. (1992) affirment que l'estimateur de Liang & Zeger (1986) donné par l'équation (1.4.7) est plus robuste que celui de Prentice (1988) donné par l'équation (1.4.8) et recommandent son utilisation lorsque le nombre de répétitions par sujet est peu élevé par rapport au nombre de sujets. Par contre Prentice (1988) insiste sur le fait que l'estimateur de la variance donné par (1.4.8) est plus stable (moins variable) que celui de Liang & Zeger (1986) donné par (1.4.7) lorsque nous sommes en présence de peu de sujets et beaucoup de répétitions. Prentice (1988) remarque que son estimateur donné par (1.4.8) subit une perte de robustesse lorsque la structure de corrélation de départ est mal spécifiée.

### 1.4.3. Structure de corrélation

La méthode des équations d'estimation généralisées repose sur le choix d'une structure de départ pour la matrice de corrélation qui reflète le plus vraisemblablement possible la structure de corrélation intra-sujet. Or, cette matrice de corrélation de départ peut être mal spécifiée sans compromettre la validité du modèle. Cela veut dire que nous pourrions utiliser la structure de corrélation de départ que nous voulons (indépendante, uniforme, auto-régressive,...), sans pour autant compromettre la consistance et la normalité asymptotique des estimateurs. Asymptotiquement ce résultat est valide mais nous allons voir qu'en pratique, nous avons quand même avantage à utiliser la vraie structure de corrélation au départ. Par exemple, si nous croyons que nos observations possèdent une structure de corrélation intra-sujet de type auto-régressive d'ordre 1 (AR(1)), nous avons probablement avantage à définir la matrice de corrélation de départ avec cette structure. Par contre, les estimateurs développés par Liang & Zeger (1986) ne tiennent pas compte de cette structure de corrélation de départ. Nous aurions

pu tout aussi bien définir cette matrice par n'importe quelle autre structure et les estimateurs des coefficients de la régression auraient encore été asymptotiquement de loi normale.

Il existe plusieurs structures de corrélation mais celles que nous utiliserons sont les suivantes.

D'abord, nous avons la structure indépendante définie par

$$R_i(\rho) = \begin{bmatrix} 1 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & 1 & 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & 1 & 0 & \cdot \\ 0 & \cdot & \cdot & \cdot & 0 & 1 \end{bmatrix}_{n_i * n_i}, \quad i = 1, \dots, s.$$

Dans cette structure, il n'y a pas de corrélation intra-sujet.

La structure de corrélation uniforme est donnée par

$$R_i(\rho) = \begin{bmatrix} 1 & \rho & \cdot & \cdot & \cdot & \rho \\ \rho & 1 & \rho & \cdot & \cdot & \rho \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & 1 & \rho & \cdot \\ \rho & \cdot & \cdot & \cdot & \rho & 1 \end{bmatrix}_{n_i * n_i}, \quad i = 1, \dots, s.$$

Cette structure suppose qu'il y a une corrélation  $\rho$  entre chaque paire de mesures pour un même sujet. Laird & Ware (1982) ont proposé l'estimateur suivant pour

$\rho$  étant donné un paramètre d'échelle  $\phi$

$$\hat{\rho} = \frac{\hat{\phi} \sum_{i=1}^s \sum_{t>t'} \hat{\gamma}_{it} \hat{\gamma}_{it'}}{\left[ \sum_{i=1}^s \frac{1}{2} n_i (n_i - 1) - p \right]}.$$

La structure de corrélation exponentielle aussi appelée structure auto-régressive pour les variables continues est définie par

$$R_i(\rho) = \begin{bmatrix} 1 & \rho & \rho^2 & \cdot & \cdot & \rho^{n_i-1} \\ \rho & 1 & \rho & \cdot & \cdot & \rho^{n_i-2} \\ \rho^2 & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & 1 & \rho & \cdot \\ \rho^{n_i-1} & \cdot & \cdot & \rho^2 & \rho & 1 \end{bmatrix}_{n_i \times n_i}, \quad i = 1, \dots, s.$$

Liang & Zeger (1986) estiment  $\rho$  par la pente de la droite de régression de  $\log(\hat{\gamma}_{it})$  sur  $\log(|t - t'|)$ .

Nous pouvons aussi imposer une corrélation nulle pour deux mesures distantes de plus de  $T$  unités de temps où  $T$  est une constante positive. La constante  $T$ , à valeur entière, fait référence à l'ordre de corrélation pour une structure auto-régressive. La structure de corrélation est alors de la forme

$$\text{corr}(y_{it}, y_{it'}) = \begin{cases} \rho^{|t-t'|} & \text{si } |t - t'| \leq T, \\ 0 & \text{sinon.} \end{cases}$$

Cette structure est appelée exponentielle par blocs.

Enfin, la structure de corrélation à effet aléatoire selon le modèle:

$$\text{logit}(P(Y_{ij} = 1|\epsilon_1)) = \text{logit}(p_{ij}) = X_{ij}\beta + \sigma\epsilon_i, \text{ où } \epsilon_i \sim N(0,1)$$

où

$$\text{logit}(A) = \log(A/(1 - A))$$

alors

$$\text{corr}(y_{ij}, y_{ik}) = \rho_{ijk} \sim \sigma^2 \sqrt{p_{ij}(1 - p_{ij})p_{ik}(1 - p_{ik})}, j \neq k,$$

où  $\rho_{ijk}$  est la corrélation intra-sujet du  $i^{\text{ième}}$  sujet entre la  $j^{\text{ième}}$  et la  $k^{\text{ième}}$  mesure (Sharples & Breslow (1992)).

#### 1.4.4. Efficacité des EEG

Liang & Zeger (1986) ont analysé les estimateurs obtenus par la méthode des EEG avec une corrélation de départ identique et différente de la vraie structure de corrélation des observations.

Avec 10 répétitions par sujet et deux niveaux de corrélation (0,30 et 0,70), ils ont calculé l'efficacité relative asymptotique (rapport des variances) entre les deux estimateurs obtenus pour toutes les combinaisons de structures de corrélation du tableau 1.2

Ce tableau contient donc le rapport des variances de l'estimateur des EEG calculé pour chacune des structures de corrélation de départ énumérées (colonnes 1 à 4) par rapport à l'estimateur des EEG calculé avec comme structure de corrélation de départ, la vraie structure de corrélation. Il est donc tout à fait normal d'obtenir 1 sur la diagonale des trois dernières colonnes de ce tableau.

TABLEAU 1.2

Efficacité relative asymptotique de  $\hat{\beta}_{EEG}$ 

Vraie corrélacion	Niveau de corrélacion	CORR. DE DÉPART			
		Indépendance	Dépendance d'ordre 1	Uniforme	AR(1)
Dépendance d'ordre 1	$\rho=0,30$	0,97	1,00	0,97	0,99
	$\rho=0,70$	0,74	1,00	0,74	0,81
Uniforme	$\rho=0,30$	0,99	0,95	1,00	0,95
	$\rho=0,70$	0,99	0,23	1,00	0,72
AR(1)	$\rho=0,30$	0,97	0,99	0,97	1,00
	$\rho=0,70$	0,88	0,75	0,88	1,00

Nous pouvons analyser ce tableau pour essayer de répondre à deux questions. Est-il vraiment utile d'utiliser la bonne structure de corrélation de départ? Et, est-ce qu'il est nécessaire d'utiliser une structure de corrélation ou ne pourrions-nous pas toujours utiliser la structure indépendante qui ne tient pas compte de la corrélation intra-sujet?

Pour répondre à la première question, il faut comparer la diagonale des trois dernières colonnes aux résultats qui ne sont pas sur la diagonale (toujours pour les trois dernières colonnes du tableau). Avec un niveau de corrélation faible ( $\rho=0,30$ ), il n'y a pas vraiment de différence. Par contre, lorsque  $\rho=0,70$ , nous remarquons des différences appréciables, surtout lorsque la structure de corrélation des observations est uniforme.

Pour répondre à la deuxième question, nous comparons la première colonne au trois dernières. Une chose est certaine, utiliser la vraie structure de corrélation de départ comme structure de corrélation de départ est toujours plus avantageux que d'utiliser la structure de corrélation indépendante. Par contre, lorsque

l'on n'utilise pas la vraie structure de corrélation des observations, aucune règle générale ne semble ressortir du tableau. Dans certains cas nous sommes mieux d'utiliser une mauvaise structure tandis que dans d'autres cas, nous serions mieux de considérer l'indépendance des observations.

A priori, il semble donc intéressant d'utiliser la bonne structure de corrélation de départ quoique Liang & Zeger (1986) ont démontré théoriquement que même si la structure de corrélation est mal spécifiée, les estimateurs suivent quand même asymptotiquement une loi multinormale.

En ce qui concerne la structure de corrélation indépendante, les résultats de Liang & Zeger (1986) ne sont pas concluants. Nous aurons l'occasion de revenir sur cette question lors de l'analyse de nos résultats de simulation.

La revue de la littérature de la section suivante nous propose des résultats beaucoup plus approfondis pour la méthode des EEG avec de petits échantillons.

## 1.5. REVUE DE LA LITTÉRATURE

Cette revue de la littérature vise à identifier les études effectuées sur la méthode des équations d'estimation généralisées appliquée à de petits échantillons. Nous avons recensé cinq articles, qui sont énumérés au tableau 1.3, traitant de ce sujet. L'intérêt principal de chaque article est très varié d'un à l'autre. Les auteurs concernés ont analysé le biais, le pourcentage de convergence des algorithmes, l'erreur de type 1, la probabilité de recouvrement et la puissance de certains tests.

Nous devons dire que même si les estimateurs obtenus avec la méthode des EEG sont différents de ceux obtenus avec les modèles linéaires généralisés, les



mêmes tests et intervalles de confiances sont utilisés pour ces deux méthodes.

**TABLEAU 1.3**  
Objectif(s) de recherche des auteurs

Auteur	Objectif(s) de recherche
Paik (1988)	Étudier l'étendue moyenne des intervalles de confiance, la probabilité de recouvrement ainsi que l'erreur de type 1 pour un test de nullité des coefficients du modèle de régression.
Emrich & coll. (1992)	Analyser l'erreur de type 1 pour le test de nullité des coefficients du modèle de régression.
Sharples & coll. (1992)	Étudier le comportement des estimateurs des EEG sous plusieurs structures de corrélation (écart type, biais, pourcentage de convergence des estimateurs et comparaison à l'estimateur du maximum de vraisemblance).
Lipsitz & coll. (1994)	Comparer la puissance, le biais et le pourcentage de convergence de trois méthodes d'estimation: la régression logistique, les EEG et les EEG après une itération seulement (EEG1).
Gunsolley & coll. (1995)	Analyser l'erreur de type 1 d'un test de proportions pour chacune des combinaisons des paramètres $\theta$ (taux de prévalence à la carie dentaire) et $\rho$ (corrélation intra-sujet).

Les auteurs nous mettent en garde avant de comparer leurs résultats à ceux des autres études à cause de la grande diversité des objectifs visés et des paramètres utilisés: nombre de sujets, nombre de répétitions par sujet, structure

de corrélation, estimateur de la variance et type de données. Les contradictions au niveau des résultats obtenus sont possibles étant donné que les auteurs ne cherchent pas tous à vérifier la même chose dans les mêmes conditions.

Dans le tableau 1.4, nous listons les paramètres utilisés pour les simulations effectuées par chaque auteur concernant les EEG avec de petits échantillons.

**TABLEAU 1.4**  
Choix des paramètres utilisés par les  
auteurs pour leurs simulations.

Auteur	Nombre de sujets	Nombre de répétitions	Structure de corrélation	Niveau de corrélation	Type de données	Estimateur de la variance
Paik (1988)	10, 20, 30 et 50	4	Uniforme, indépendante et AR(1)	0,30 et 0,40	Gamma	Liang & Zeger (1986)
Emrich & coll. (1992)	20	64	Indépendante, exponentielle et expo. par blocs	0,30	Binaire	Liang & Zeger (1986) et Prentice (1988)
Sharples & coll. (1992)	80, 100 et 200	2, 4 et 6	Uniforme, indépendante, effet aléatoire et AR(1)	0,10 et 0,30	Binaire	Liang & Zeger (1986) et Prentice (1988)
Lipsitz & coll. (1994)	15, 30 et 45	3	Uniforme	0,15, 0,45 et 0,60	Binaire	Liang & Zeger (1986)
Gunsolley & coll. (1995)	20	100	Uniforme	0,10, 0,20 et 0,30	Beta-binômiale	Liang & Zeger (1986) et Prentice (1988)

L'article de Gunsolley & coll. (1995) utilise des observations sans aucune dépendance temporelle. Nous sommes en présence de variables corrélées mesurées simultanément contrairement aux autres études où les répétitions sont mesurées à des temps différents. Les matrices de corrélation de départ pour chaque structure de corrélation citée dans ce tableau sont décrites théoriquement à la section 1.4.3 .

Le tableau 1.5 résume les résultats obtenus par les trois auteurs qui ont utilisé des données binaires.

**TABLEAU 1.5**

Comparaison des résultats pour des données binaires avec l'estimateur de la variance de Liang & Zeger (1986) et l'estimateur de Prentice (1988).

Auteur	Paramètres	Résultats
Emrich & coll. (1992)	20 sujets et 64 répétitions par sujet	Un mauvais choix de la structure de corrélation de départ peut impliquer une sous-estimation ou une surestimation de l'estimateur de la variance selon la structure de corrélation utilisée. Les auteurs suggèrent l'estimateur de Prentice (1988) lorsque la structure de corrélation de départ est bien spécifiée ou lorsque toutes les structures de corrélations plausibles mènent aux mêmes conclusions. Sinon, il faut utiliser l'estimateur de la variance de Liang & Zeger (1986).
Sharples & coll. (1992)	80 à 200 sujets et 2 à 6 répétitions par sujet	Il n'y a aucune différence avec les deux estimateurs de la variance mais ils reconnaissent qu'avec moins de sujets, ils auraient probablement trouvé des résultats plus discriminants.
Gunsolley & coll. (1995)	20 sujets et 100 répétitions par sujet	L'estimateur de la variance de Liang & Zeger (1986) surestime toujours l'erreur de type 1 tandis que l'estimateur de la variance de Prentice (1988) sous-estime l'erreur de type 1 mais à un moindre niveau.

Comparons maintenant, dans le tableau 1.6, les résultats des quatre auteurs qui ont analysé le biais des estimateurs de la régression avec la méthode des EEG.

Nous pouvons inférer que pour obtenir des coefficients non biaisés, il faut avoir plusieurs répétitions par sujet (Gunsolley & coll. (1995) et Emrich & coll. (1992)) à défaut de quoi il faut un très grand nombre de sujets (Sharples & coll. (1992)).

**TABLEAU 1.6**  
 Comparaison des résultats concernant  
 le biais des estimateurs.

Auteur	Résultats
Paik (1988)	<p>Les estimateurs sont biaisés de l'ordre de 18% à 109% avec 10 sujets et de l'ordre de 2% à 21% avec 50 sujets.</p> <p>Trois types de jackknife (ordinaire, pondéré et linéaire) sont proposés comme méthode de réduction du biais.</p>
Emrich & coll. (1992)	<p>Les EEG produisent des estimateurs sans biais même avec de petits échantillons indépendamment du choix de la matrice de corrélation de départ.</p>
Sharples & coll. (1992)	<p>Le biais maximum est de 9% avec un coefficient de corrélation de 0,10 et de 10,5% avec un coefficient de corrélation de 0,30.</p> <p>Sans preuve à l'appui, les auteurs affirment que doubler le nombre de sujets donne de meilleurs résultats, en ce qui concerne le biais, que de doubler le nombre de répétitions pour chaque sujet et qu'avec 200 observations l'indépendance pourrait être utilisée comme matrice de corrélation de départ, c'est-à-dire ne pas tenir compte de la corrélation intra-sujet.</p>
Lipsitz & coll. (1994)	<p>Il n'y a aucune différence entre les trois méthodes (la régression logistique, la méthode des EEG avec seulement une itération et la méthode des EEG ordinaire) au niveau du biais des estimateurs. Les auteurs n'ont pas commenté les biais obtenus par les trois méthodes pré-citées.</p>
Gunsolley & coll. (1995)	<p>Le biais des coefficients de la régression est inférieur à 0,20 .</p>

Nous analyserons maintenant (tableau 1.7) les taux de convergence des algorithmes utilisés pour résoudre les équations d'estimation généralisées. L'article de

Lipsitz & coll. (1994) est particulièrement intéressant parce qu'il offre une solution aux algorithmes non convergents.

**TABLEAU 1.7**

Convergence des algorithmes.

Auteur	Résultats
Lipsitz & coll. (1994)	<p>Les plus sérieux problèmes de convergence des algorithmes sont survenus avec une forte corrélation intra-sujet (<math>\rho=0,60</math>) et peu de sujets (15).</p> <p>Ces problèmes sont dûs à la singularité de la matrice de covariances ou à la non convergence de <math>\hat{\beta}</math>.</p> <p>Lorsque les estimateurs n'ont pas convergé, une comparaison des estimés des EEG après seulement une itération (EEG1) nous montre que les estimateurs des EEG sont bons après seulement une itération.</p>

Une partie importante des études de simulation appliquées aux petits échantillons avec la méthode des EEG est l'analyse de l'erreur de type 1 d'un test ou la probabilité de recouvrement d'un intervalle de confiance, qui estime respectivement  $\alpha$  et  $1-\alpha$ . Dans le tableau 1.8, nous résumons les résultats concernant l'erreur de type 1.

**TABLEAU 1.8**  
Erreur de type 1.

Auteur	Résultats
Emrich & coll. (1992)	L'estimateur de Liang & Zeger (1986) donne un test global pour le modèle de régression ( $\chi_4^2$ ) et des tests bilatéraux sur chacun des 4 coefficients du modèle (test t) dont les niveaux sont plus grands que le 5% spécifié. L'estimé de la variance de Prentice (1988) ne produit pas de bons résultats en ce qui concerne l'erreur de type 1, à moins d'utiliser la bonne structure de corrélation de départ. Par contre, l'estimateur de Liang & Zeger (1986) produit des estimés de l'erreur de type 1 plus constants indépendamment de la structure de corrélation de départ.
Gunsolley & coll. (1995)	Avec 20 sujets et 100 répétitions, l'estimateur de la variance de Liang & Zeger (1986) donne une erreur de type 1 variant entre 5,8% et 25,3% pour le test de Wald( $\alpha=0,05$ ). Des résultats très légèrement meilleurs sont obtenus avec l'estimateur de la variance de Prentice (1988) avec une erreur de type 1 variant entre 0,69% et 6,2%. Nous remarquons une surestimation de l'erreur de type 1 avec l'estimé de la variance de Liang & Zeger (1986) contrairement à la sous-estimation de l'erreur de type 1 avec l'estimé de la variance de Prentice (1988).

De ces deux articles, nous pouvons possiblement avancer que l'estimateur de la variance de Liang & Zeger (1986) est porté à surestimer l'erreur de type 1 tandis que l'estimateur de la variance de Prentice (1988) a tendance à la sous-estimer.

Paik (1988) a porté une attention particulière à l'étendue des intervalles de confiance avec les estimateurs des EEG. Il a remarqué que l'étendue moyenne des intervalles avec 20 sujets sous une structure de corrélation de départ bien spécifiée (AR(1)) est aussi grande que celle avec 50 sujets sous une structure de corrélation indépendante. C'est donc dire que le fait de bien spécifier la structure de corrélation plutôt que de spécifier l'indépendance, nous permet d'obtenir sensiblement



les mêmes résultats avec seulement 40% de l'échantillon.

Seul Emrich & coll. (1992) nous ont fait part de leurs résultats au sujet de la puissance de leurs tests. La puissance du test global et des tests bilatéraux sur chacun des coefficients, est supérieure à 92% dans tous les cas sauf pour un de ces quatre coefficients. L'estimateur de Liang & Zeger (1986) donne une puissance de test stable, c'est-à-dire constante en fonction des différentes structures de corrélation de départ. Ce qui n'est pas nécessairement le cas avec l'estimateur de Prentice (1988) qui nous donne une puissance avec un peu plus de variations.

### 1.5.1. Conclusion

La majorité des résultats tirés de ces cinq articles ne sont pas comparables. En ce qui concerne les statistiques traitées par plus d'un article (biais, erreur de type 1 et estimateur de la variance), nous trouvons des contradictions au niveau des résultats étant donné les différences dans les autres paramètres utilisés pour effectuer ces analyses.

Ces différents résultats n'ont pas été établis d'un point de vue général ou comparatif. Ils sont plutôt indicatifs et partiels. Le plus bel exemple est que quatre des cinq articles ont utilisé un nombre de sujets et/ou de répétitions par sujet fixe pour toutes leurs analyses. Or, il s'agit pourtant de deux paramètres très importants à faire varier lorsque l'on étudie les propriétés d'une méthode avec de petits échantillons.

Cette revue de la littérature n'aura donc servi qu'à présenter une partie des résultats obtenus. Par contre, nous pourrons mieux les commenter après avoir effectué les analyses proposées au chapitre suivant. Nous pourrons alors, lorsque les paramètres utilisés seront comparables, comparer nos résultats avec ceux déjà

obtenus.

# Chapitre 2

---

## ANALYSE DES RÉSULTATS

### 2.1. LA SIMULATION

A la suite de la revue de la littérature présentée dans le chapitre précédent, il s'avère que pour mieux cerner la pertinence d'utiliser la méthode des équations d'estimation généralisées avec un petit échantillon, il est utile de poursuivre la comparaison des estimateurs de la variance de Liang & Zeger (1986) et de Prentice (1988). Nous analyserons aussi les coefficients de la régression (biais, minimum, maximum, étendue et écart type).

Pour ce faire nous avons effectué des simulations en faisant varier une série de paramètres: le nombre de sujets, le nombre de répétitions par sujet, la structure de corrélation des données générées, la structure de corrélation de départ, le type de données et l'estimateur de la variance. Dans le tableau 2.1, nous retrouvons la liste exhaustive de tous les paramètres que nous avons fait varier lors des simulations.

**TABLEAU 2.1**  
Paramètres des simulations

Structure de corrélation des observations	Uniforme et AR(1)	
Structure de corrélation de départ	Uniforme, AR(1) et indépendante	
Estimateur de la variance	Liang & Zeger (1986) et Prentice (1988)	
Nombre de sujets	20 et 100	
Nombre de répétitions par sujet	5 et 10	
Type de données	normale	binaire
Niveau de corrélation	0,15 et 0,45	0,15 et 0,30

Nous n'avons pas utilisé les mêmes niveaux de corrélation pour les deux types de données. Cela est dû au haut taux de non-convergence de l'algorithme avec des données binaires.

Pour toutes les combinaisons de ces paramètres, nous avons effectué 2500 répétitions dans le cas des données normales et 1000 pour les données de type binaire. Nous avons été obligés de réduire le nombre de répétitions par sujet pour les données de type binaire puisque le principe de dichotomisation pour obtenir

des données de type binaire avec une corrélation intra-sujet non-nulle est extrêmement long en temps-machine.

Toutes les simulations ont été effectuées à l'aide de trois ordinateurs: portable IBM 380 Pentium 150 megahertz avec 16 mégaoctets de mémoire vive, Hewlett Packard (Vectra) Pentium 200 megahertz avec une mémoire vive de 32 mégaoctets ainsi qu'un COMPAQ (DESKPRO) Pentium Pro 200 megahertz avec 64 mégaoctets de mémoire vive.

En ce qui concerne la génération des données de type normal, nous avons considéré un modèle de régression multiple avec 4 variables indépendantes. Étant donné que la matrice des 4 variables indépendantes ( $X$ ) et le vecteur des 5 coefficients de la régression ( $\beta$ ) sont fixes, cela implique que la structure de corrélation intra-sujet du vecteur d'observations de la variable dépendante ( $\underline{Y}$ ) est la même que celle du vecteur d'erreur  $\underline{\epsilon}$ .

Pour construire la matrice ( $M$ ) avec la structure de corrélation que nous souhaitons imposer à la variable dépendante du modèle, nous nous sommes servis de

$$\begin{aligned}\epsilon &\sim N(0, \sigma^2) \\ M'\epsilon &\sim N(0, M'\sigma^2 M) \\ Y = X\beta + M'\epsilon &\sim N(X\beta, M'\sigma^2 M)\end{aligned}$$

Quant aux valeurs du terme d'erreur  $\epsilon$ , elles ont été générées à partir du générateur de nombres pseudo-aléatoires du logiciel SAS (1990). Chaque  $\epsilon$  est égale à une valeur générée selon une variable de loi normale de moyenne 0 et de variance 25. Le germe de départ ainsi que le pas ont été fixés à 5000. Nous avons donc réussi à obtenir des valeurs pour une variable dépendante ( $Y$ ) possédant une

structure de corrélation que nous contrôlons.

En ce qui concerne les valeurs de la variable dépendante de type binaire, nous avons utilisé la méthode de Emrich, Marion & Piedmonte (1991) pour transformer des observations multinormales avec une certaine structure de corrélation en observations de type binaire tout en conservant la structure de corrélation. Cette méthode, qui est relativement complexe, utilise la fonction de distribution cumulative d'un vecteur aléatoire bivarié de type normal avec un coefficient de corrélation  $\rho$  pour résoudre une équation avec la méthode de la bisection. Les valeurs ainsi obtenues nous servent à générer des données provenant d'une variable aléatoire multivariée de loi normale. Ces données ainsi obtenues seront dichotomisées en fonction qu'elles sont plus grandes ou non qu'un  $p^e$  quantile ( $p$  varie pour chaque observation) d'une variable aléatoire  $N(0,1)$ . Pour plus de précision, nous pouvons nous référer à l'article ainsi qu'à l'algorithme (annexe A) et les programmes (annexe B).

Nous avons choisi des valeurs différentes pour les variables indépendantes ( $X$ ) et les coefficients de la régression ( $\beta$ ). Ces changements ont été rendus nécessaires après quelques problèmes dus aux traitements de ces observations pour obtenir des données binaires (log(0), matrice singulière, ...). Nous avons utilisé le même vecteur  $\underline{\beta}$  qu'Emrich & Piedmonte (1992) mais cela n'était pas dans le but de comparer directement nos résultats aux leurs puisque les paramètres primaires utilisés pour nos simulations (nombre de sujets, nombre de répétitions par sujet et les structures de corrélation) sont très différents des leurs.

Nous avons choisi d'utiliser la macro de Karim & Zeger (1988) parce qu'elle utilisait un logiciel (SAS) très performant quant à la gestion des données et à

l'extraction des résultats, quoique relativement lente lors de l'exécution des simulations. Les versions Fortran, Sudaan et Splus ne nous offraient pas cette flexibilité. Nous avons passablement modifié cette macro pour faciliter la gestion des résultats. Cette macro SAS modifiée ainsi que toutes les autres macros que nous avons construites pour générer les données et les traiter sont présentées à l'annexe B. La version 6.12 de SAS inclut maintenant une procédure (GENMOD) pour faire ce type d'analyse. Cette procédure requiert moins de temps-machine que la macro utilisée puisqu'elle a été optimisée et qu'elle est programmée en langage C (langage de programmation de SAS) contrairement à la macro où tous les résultats sont obtenus à partir de programmes écrits directement en SAS. Quelques mois seulement ont séparé le début de mes simulations et la sortie de cette nouvelle version de SAS.

Après avoir décrit de quelle manière les données ont été générées, il est maintenant temps de passer à la plus importante des sections de ce mémoire: l'analyse des résultats.

## 2.2. ANALYSE DES RÉSULTATS

Les 24 tableaux de résultats (tableaux 3.1 à 8.4) présentent, pour une série de paramètres fixés, les résultats obtenus pour chaque simulation.

Les statistiques que nous avons retenues pour analyser les coefficients de la régression sont: l'estimé de chaque coefficient, l'écart type, le minimum, le maximum, l'étendue, le biais relatif, ainsi que l'estimé de la variance selon Liang & Zeger (1986) et Prentice (1988). Tous les estimés de ces statistiques sont en fait une moyenne obtenue de toutes les répétitions (2500 pour les données normales et 1000 pour les données binaires) des simulations.

Nous analyserons chacune de ces statistiques en fonction des différents paramètres utilisés. À première vue, les résultats obtenus semblent passablement différents en ce qui concerne les deux types de données (normales et binaires). Étant donné qu'aucun des articles traités dans notre revue de la littérature compare les deux types de données, nous les analyserons séparément. Pour la suite de ce chapitre, nous allons nous concentrer à analyser le coefficient  $\beta_0$ . Aucune caractéristique ne différencie  $\beta_0$  des autres coefficients ( $\beta_1, \dots, \beta_4$ ). Nous ne devons pas nous attendre à des résultats différents de la part de ces coefficients. Dans le seul but d'éviter de nous répéter à rapporter des résultats équivalents, nous ne rapporterons que les résultats du coefficient  $\beta_0$  à moins que cela soit pertinent. Les tableaux de résultats 3.1 à 8.4 contiennent les résultats de tous les coefficients.

Traisons d'abord les données de type normal. Les valeurs moyennes obtenues pour  $\beta_0$  varient entre 1,96 et 2,04 pour un biais relatif variant entre -0,02 et 0,02 ( $\beta_0 = 2$ ). Nous pouvons donc affirmer que les coefficients de la régression obtenus avec des données normales ne sont pas biaisés et ce de manière indépendante en ce qui concerne le nombre de sujets, le nombre de répétitions par sujet, la structure de corrélation des observations et de départ ainsi que le niveau de corrélation.

Logiquement, l'étendue (maximum - minimum) du coefficient  $\beta_0$  diminue lorsque l'on passe de 5 à 10 répétitions par sujet avec 20 et 100 sujets et ce pour toutes les combinaisons de paramètres. Cependant, nous remarquons que l'étendue de ce coefficient est toujours plus élevée avec 100 sujets et 5 répétitions par sujet qu'avec 20 sujets et 10 répétitions par sujet. Or, Sharples & coll. (1992) avaient affirmé qu'il était préférable de doubler le nombre de répétitions par sujet que le nombre de sujets. En ce qui concerne l'étendue lors de nos simulations, cette affirmation est toujours vraie même lorsque l'on quintuple (20 à 100) le nombre de sujets versus doubler le nombre de répétitions par sujet (5 à 10). Nous avons obtenu la plus petite étendue avec 20 sujets et 10 répétitions par sujet qui est, en



moyenne, deux fois plus petite que celle obtenue avec 100 sujets et 5 répétitions par sujet.

Quant aux niveaux de corrélation intra-sujet (0,15 et 0,45), un coefficient de corrélation élevé (0,45) donne de plus petites étendues 22 fois sur 24. La variabilité des observations est moins élevée avec un coefficient de corrélation intra-sujet plus élevé. Au niveau des structures de corrélation, nous n'observons pas d'étendue plus grande lorsque nous spécifions mal la structure de corrélation de départ, c'est-à-dire lorsqu'elle est différente de la structure de corrélation que nous avons utilisée pour générer les données. Avant de conclure, nous analyserons les différentes structures de corrélation en fonction des autres statistiques mais a priori le fait de mal spécifier la structure de corrélation de départ ne semble pas affecter les résultats même avec de petits échantillons.

Puisque nous avons obtenu un biais relatif nul, nous n'avons pas calculé l'écart quadratique moyen ( $\text{écart type}^2 + \text{biais}^2$ ), nous allons seulement analyser l'écart type. Le même principe se reproduit en ce qui concerne l'écart type de la moyenne des 2500 coefficients de la régression pour chaque simulation. Nous avons donc avantage à augmenter le nombre de répétitions par sujet plutôt que le nombre de sujets. Aucune différence n'est ressortie concernant les structures de corrélation sauf qu'avec un niveau de corrélation élevé (0,45) on obtient un plus petit écart type 20 fois sur 24 comparativement à un niveau de corrélation moins élevé (0,15). Ces 4 cas sont survenus lorsque la structure de corrélation des observations était de type AR(1).

Analysons maintenant les deux estimateurs de l'écart type du coefficient de la régression  $\beta_0$ . Rappelons que l'estimateur de la variance de Prentice (1988) a été spécifiquement développé pour des données de type binaire. Les articles traités dans la revue de la littérature qui ont utilisé l'estimateur de Prentice (1988)

utilisaient tous des données de type binaire. Nous comparerons donc l'estimateur de Prentice (1988) à celui de Liang & Zeger (1986) mais seulement d'une manière exploratoire.

Pour des données de loi normale, indépendamment de tous les paramètres, l'estimateur de l'écart type de Liang & Zeger (1986) sous-estime toujours l'écart type de  $\beta_0$  obtenu à partir des simulations. L'estimateur de l'écart type de Prentice (1988) sous-estime l'estimateur de l'écart type de  $\beta_0$  pour 26 choix de paramètres sur 48 possibilités. Rien de particulier semble caractériser les deux estimés entre eux. L'estimé de Prentice (1988) est plus élevé que celui de Liang & Zeger (1986) pour 28 simulations sur 48.

L'estimateur de la variance de Prentice (1988) est toujours plus élevé avec un niveau de corrélation intra-sujet moins élevé (0,15) qu'avec un niveau élevé (0,45). Par contre, l'estimateur de Liang & Zeger (1986) est plus élevé 20 fois sur 24 avec un niveau moins élevé de corrélation intra-sujet (0,15) comparativement à un niveau élevé (0,45).

Traitions maintenant des résultats avec des données de type binaire. Une variable dépendante de type normal contient beaucoup plus d'information qu'une autre de type binaire. Il est donc prévisible que les résultats obtenus avec des données binaires soient beaucoup plus variables que les résultats précédemment analysés. Comme pour les données normales, nous analyserons seulement les résultats du coefficient  $\beta_0$  dans la mesure où ils ne diffèrent pas des résultats obtenus avec les autres coefficients.

A cause du problème énoncé auparavant, les paramètres utilisés pour les simulations sont différents pour les données de type binaire en ce qui concerne le

niveau de corrélation. Les deux niveaux de corrélation utilisés seront 0,15 et 0,30 .

Les résultats obtenus nous montrent un biais variant entre -0,03 et 0,13 pour le coefficient  $\beta_0$ . Nous en concluons donc, comme avec des données de type normal, à une absence de biais indépendamment des paramètres de simulation. Il est intéressant de savoir qu'avec seulement 5 répétitions par sujet nous obtenons des estimateurs non-biaisés. Les coefficients  $\beta_1, \dots, \beta_4$  sont aussi sans biais.

Des résultats encore plus intéressants sont survenus à l'analyse de l'étendue du coefficient  $\beta_0$ . Le Tableau 2.2 nous indique les résultats obtenus. Nous ne remarquons pas vraiment de différence en ce qui concerne la structure de corrélation des observations et de départ ainsi que le niveau de corrélation (0,15 et 0,30). Par contre, nous avons obtenu des étendues moyennes avec 5 répétitions par sujet de 29,35 et 3,18 respectivement pour 20 et 100 sujets et des étendues moyennes de 2,85 et 0,79 pour 20 et 100 sujets avec 10 répétitions par sujet. L'étendue est entre 3 et 10 fois plus petite lorsque nous passons de 20 à 100 sujets et dans les mêmes proportions lorsque nous passons de 5 à 10 répétitions par sujet. Cela nous semble correct étant donné que nous utilisons 5 fois plus de sujets (20 versus 100). Cependant, cette proportion semble très importante en ce qui concerne le nombre de répétitions par sujet. Rappelons que nous avons seulement doublé le nombre de répétitions par sujet pour obtenir ces proportions. Passer d'une étendue moyenne de 29,35 à 2,85 en faisant seulement varier le nombre de répétitions par sujet de 5 à 10 n'est pas négligeable.

Par ordre décroissant d'importance, nous avons obtenu les meilleurs résultats, en ce qui concerne l'étendue moyenne, avec beaucoup de répétitions par sujet (10 versus 5), beaucoup de sujets (100 versus 20), un niveau de corrélation faible (0,15

versus 0,30), une structure de corrélation des observations de type AR(1) (plutôt qu'uniforme) et ce indépendamment de la structure de corrélation de départ.

**TABLEAU 2.2**

Étendue du coefficient  $\beta_0$   
(données binaires)

Niveau de corrélation		0,15				0,30			
		5		10		5		10	
		20	100	20	100	20	100	20	100
Struc. de corrélation des observations	Struc. de corrélation de départ								
Uniforme	Uniforme	60,35	3,04	2,82	0,80	24,6	3,57	3,93	0,80
	AR(1)	20,34	3,04	3,01	0,80	24,6	3,57	4,09	0,86
	Indépendante	60,34	3,04	2,81	0,80	24,6	3,57	3,37	0,83
AR(1)	Uniforme	18,54	2,80	2,16	0,73	53,43	3,32	2,52	0,80
	AR(1)	18,53	2,80	2,36	0,66	14,18	3,31	2,50	0,83
	Indépendante	18,53	2,80	2,16	0,73	14,18	3,31	2,51	0,80

Puisque les résultats obtenus avec des données de type binaire sont non-biaisés, il n'est pas intéressant d'analyser l'écart quadratique moyen. Cela revient à analyser la variance dont les résultats sont donnés dans le tableau 2.3.

Nous voici arrivés aux dernières analyses: l'analyse des deux estimateurs de l'écart type (Liang & Zeger (1986) et Prentice (1988)).

A première vue, nous ne remarquons pas de différence majeure entre l'estimateur de Liang & Zeger (1986) et celui de Prentice (1988) pour un même choix de paramètres. Si nous comparons ces deux estimateurs séparément en faisant varier un paramètre, c'est le nombre de répétitions par sujet qui nous offre le plus de variation avec un rapport d'environ 4 pour 1 si nous comparons les résultats obtenus avec 5 et 10 répétitions par sujet. Si nous faisons varier le nombre de sujets, nous obtenons un rapport d'environ 2 pour 1 avec 20 et 100 sujets respectivement. Les autres paramètres: structure de corrélation des observations et de départ ainsi que le niveau de corrélation n'influencent pas vraiment les résultats. La comparaison de ces deux estimateurs à celui obtenu lors des simulations nous indiquera un peu mieux le comportement de ces estimateurs. Avec 100 sujets, les deux estimateurs surestiment toujours celui des simulations et cela pour tous les autres paramètres. L'estimateur de Prentice (1988) est plus variable puisqu'avec 20 sujets, il sous et surestime l'estimateur des simulations de manière aléatoire. Avec 20 sujets et 5 répétitions par sujet, l'estimateur de Liang & Zeger (1986) sous-estime toujours l'estimé des simulations alors qu'avec 20 sujets et 10 répétitions par sujet nous remarquons une surestimation presque à tout coup.

Les valeurs obtenues à partir de l'estimateur de l'écart type des simulations de  $\beta_0$  avec 20 sujets sont très petites, les estimateurs de Liang & Zeger (1996) et Prentice (1988) les surestiment donc toujours. Ce phénomène ne se produit pas avec 100 sujets puisque les estimateurs de Liang & Zeger (1996) et Prentice (1988) varient autour de la valeur la plus élevée de l'estimateur des simulations. Sinon, nous aurions pu penser que les estimateurs de Liang & Zeger (1986) et de Prentice (1988) sont moins bons avec beaucoup de sujets (100) qu'avec peu de sujets (20).

A la lumière de ces résultats, nous remarquons pas de différence entre les 2 estimateurs de Liang & Zeger (1986) et Prentice (1988) lorsque l'on utilise 100

sujets ou 20 sujets et 10 répétitions par sujet. Par contre, l'estimateur de Liang & Zeger (1986) possède un avantage certain sur celui de Prentice (1988) avec 20 sujets et 5 répétitions par sujet.

**TABLEAU 2.3**  
Estimateur de l'écart type du coefficient  $\beta_0$   
(données binaires)

Niveau corrélation			0,15				0,30			
Nombre de répétitions			5		10		5		10	
Nombre de sujets			20	100	20	100	20	100	20	100
Struc. corrélation des observations	Struc. corr. de départ	Estimateur de la variance								
Uniforme	Uniforme	Simulation	3,94	0,49	0,42	0,11	2,94	0,57	0,57	0,11
		Liang & Zeger	2,24	0,89	0,45	0,20	2,04	0,81	0,46	0,20
		Prentice	44,04	0,91	0,46	0,20	9,25	0,81	0,46	0,20
	AR(1)	Simulation	3,32	0,49	0,47	0,11	2,94	0,57	0,65	0,13
		Liang & Zeger	2,18	0,89	0,45	0,20	2,03	0,81	0,47	0,21
		Prentice	2,12	0,86	0,46	0,20	9,12	0,76	0,45	0,19
	Indép.	Simulation	3,66	0,49	0,42	0,11	2,94	0,57	0,57	0,12
		Liang & Zeger	2,23	0,89	0,45	0,20	2,04	0,81	0,46	0,20
		Prentice	23,29	0,96	0,46	0,20	9,57	0,96	0,47	0,20
AR(1)	Uniforme	Simulation	3,09	0,41	0,31	0,10	2,75	0,48	0,42	0,13
		Liang & Zeger	2,07	0,84	0,46	0,20	1,84	0,70	0,47	0,21
		Prentice	2,29	0,94	0,45	0,20	22,03	0,91	0,46	0,20
	AR(1)	Simulation	3,09	0,41	0,33	0,11	2,35	0,48	0,40	0,13
		Liang & Zeger	2,07	0,84	0,45	0,20	1,80	0,70	0,44	0,20
		Prentice	2,12	0,86	0,45	0,20	1,90	0,76	0,44	0,20
	Indép.	Simulation	3,09	0,41	0,31	0,10	2,34	0,48	0,43	0,13
		Liang & Zeger	2,07	0,84	0,46	0,20	1,80	0,70	0,47	0,21
		Prentice	2,35	0,96	0,46	0,20	2,36	0,96	0,46	0,20

### 2.3. CONCLUSION

Les conclusions faites dans cette section s'appliquent nécessairement qu'aux résultats obtenus avec les paramètres utilisés. Il serait dangereux de généraliser ces résultats lorsque les paramètres utilisés pour simuler ces résultats sont différents.

Quelques points importants sont apparus clairement lors de l'analyse des résultats. Nous avons remarqué une absence de biais pour des données normales et binaires. Aucun des articles analysés dans la revue de la littérature traitait d'analyse faite avec des données normales mais nous contredisons les résultats (données binaires) de Sharples & coll. (1992) et Lipsitz & coll. (1994) qui ont obtenu des estimateurs biaisés. Ces résultats ont été obtenus avec de 2 à 6 répétitions par sujet. Nous appuyons les résultats de Emrich & coll. (1992) puisqu'ils ont obtenu des estimateurs non biaisés. Cependant, ils ont utilisé 64 répétitions par sujet ce qui est beaucoup plus que ce que nous avons utilisé (10).

Nous pouvons inférer que pour obtenir des coefficients non biaisés, il faut avoir plusieurs répétitions par sujet (Gunsolley & coll. (1995) et Emrich & coll. (1992)) à défaut de quoi il faut un très grand nombre de sujets (Sharples & coll. (1992)). Cela vient appuyer l'affirmation de Sharples & coll. (1992) qui disaient qu'il y avait plus d'avantages à augmenter le nombre de répétitions par sujet que le nombre de sujets (données normales). Lorsque le nombre de répétitions par sujet est fixe (données binaires), il y a quand même un avantage certain à augmenter le nombre de sujets.

Un autre fait est apparu clairement. Nous n'avons pas obtenu de moins bons résultats lorsque la structure de corrélation de départ était mal spécifiée. Liang & Zeger (1986) avaient donc raison de dire que leur méthode produisait des estimateurs qui suivent quand même asymptotiquement une loi multinormale et ce

indépendamment de la structure de corrélation de départ.

A quelques reprises, nous avons obtenu de meilleurs résultats lorsque la structure de corrélation des observations était de type AR(1) plutôt que de type uniforme. Le problème est que la structure de corrélation des observations est inconnue en pratique. Et, comme ce fait s'est produit indépendamment de la structure de corrélation de départ et du niveau de corrélation des observations, cela est donc un résultat moins intéressant.

Le même principe s'est produit avec le niveau de corrélation. Ce paramètre n'est évidemment pas contrôlé en pratique. Nous avons obtenu de plus petites étendues et de plus petits écarts types avec un niveau de corrélation faible (0,15) (données binaires) contrairement aux données normales où cela s'est produit avec un niveau de corrélation élevé (0,45). La corrélation intra-sujet est bel et bien une nuisance lors de l'estimation des différents estimateurs.

Le paramètre clé de l'analyse longitudinale par la méthode des équations d'estimation généralisées est définitivement le nombre de répétitions par sujet. A tous les niveaux, nous avons remarqué une différence appréciable des résultats lorsque nous utilisons 10 répétitions par sujet au lieu de seulement 5. Nous avons toujours remarqué un meilleur ratio d'amélioration lorsque nous doublons le nombre de répétitions par sujet plutôt que lorsque nous quintuplions le nombre de sujets de 20 à 100. Dans la majorité des cas, nous avons obtenu des résultats très similaires avec 20 sujets et 10 répétitions par sujet qu'avec 100 sujets et 5 répétitions par sujet. Cela revient donc à dire que nous avons obtenu d'aussi bons résultats avec seulement 40% de la taille d'échantillon du second groupe (20X10 versus 100X5). Cela n'est pas négligeable lorsqu'un coût très élevé est associé à chaque donnée à recueillir.



Aucun des deux estimateurs de la variance ne s'est démarqué. Les deux donnent des résultats très semblables à quelques exceptions près.

Il nous reste donc à tirer une conclusion sur la question que nous nous posons depuis le début. L'analyse à l'aide de la méthode des équations d'estimation généralisées est-elle appropriée pour de petits échantillons? Même si la définition de petits échantillons est très relative, il est possible d'utiliser ce type d'analyse avec peu de données mais en étant très prudents et conscients que nous avons grandement avantage à augmenter le nombre de répétitions par sujet même au détriment du nombre de sujets.

# Annexe A

---

## DONNÉES BINAIRES CORRÉLÉES

Nous allons décrire la méthode utilisée pour générer des données binaire corrélés tel qu'écrit dans Emrich & Piedmonte (1991).

Le problème que nous avons est de générer  $J$  variables binaire  $Y_1, \dots, Y_J$  tel que  $E(Y_j) = p_j$  et  $\text{corr}(Y_j, Y_k) = \delta_{jk}$  où  $p_j, j = 1, \dots, J$  et  $\delta_{jk}, j = 1, \dots, J - 1; k = 2, \dots, J$  sont connus.

Soit

$$\phi(x_1, x_2, \rho) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} f(z_1, z_2, \rho) dz_1 dz_2 \quad (\text{A.0.1})$$

la fonction de distribution cumulative d'un vecteur aléatoire bivarié avec un coefficient de corrélation  $\rho$  où

$$f(z_1, z_2, \rho) = [2\pi(1 - \rho^2)^{1/2}]^{-1} \exp[-(z_1^2 - 2\rho z_1 z_2 + z_2^2)/2(1 - \rho^2)].$$

$\phi(x_1, x_2, \rho)$  est une fonction strictement croissante pour  $\rho$  étant donné  $x_1$  et  $x_2$ .

La première étape consiste à résoudre l'équation

$$\phi(z(p_j), z(p_k), \rho_{jk}) = \delta_{jk} (p_j q_j p_k q_k)^{1/2} + p_j p_k \quad (\text{A.0.2})$$

pour  $\delta_{jk} (j = 1, \dots, J - 1; k = 2, \dots, J)$  où  $z(p)$  est le  $p^e$  quantile d'une variable aléatoire normale centré et réduite.

Nous avons utilisé la méthode de la bisection et la fonction PROBBNRM de SAS qui évalue la fonction  $\phi()$  pour résoudre l'équation (A.0.1) pour  $\rho_{jk}$ .

La deuxième étape consiste à générer  $Z = (Z_1, \dots, Z_J)^t$  suivant un vecteur aléatoire multinormal de dimension  $J$ , de moyenne  $\underline{Q}$  et de matrice de corrélation  $\sigma = (\rho_{jk})$ . Pour ce faire, nous avons utilisé la macro MNV de SAS.

Finalement, pour  $j=1, \dots, J$ , soit  $Y_j = 1$  si  $Z_j \leq z(p_j)$  et  $Y_j = 0$  sinon.

Le vecteur  $Y = (Y_1, \dots, Y_J)$  possède donc les propriétés désirées puisque

$$E(Y_j) = P(Y_j = 1) = P(Z_j \leq z(p_j))$$

et

$$\begin{aligned} \text{corr}(Y_j, Y_k) &= \text{cov}(Y_j, Y_k) / (p_j q_j p_k q_k)^{1/2} \\ &= (P(Y_j = 1, Y_k = 1) - p_j p_k) / (p_j q_j p_k q_k)^{1/2} \\ &= (P(Z_j \leq z(p_j), Z_k \leq z(p_k)) - p_j p_k) / (p_j q_j p_k q_k)^{1/2} \\ &= [\phi[z(p_j), z(p_k), \rho_{jk}] - p_j p_k] / (p_j q_j p_k q_k)^{1/2} \\ &= \delta_{jk} \end{aligned}$$

par l'équation (A.0.2).

## Annexe B

---

### TABLEAUX DE RESULTATS

TABLEAU 3.1

Corrélation des observations: Uniforme

Corrélation de départ: Uniforme

Niveau de corrélation: 0,15

Type de données: Normale

	# de sujets:=20		# de sujets:=100	
	# de répétitions par sujet:=5	# de répétitions par sujet:=10	# de répétitions par sujet:=5	# de répétitions par sujet:=10
beta0 (min,max)	2,04 (-42,09;52,20)	1,98 (-11,19;12,15)	1,96 (-18,36;24,13)	2,01 (-2,66;6,21)
biais relatif	0,02	-0,01	-0,02	0,01
écart type <sup>1</sup>	13,93	3,01	6,29	1,39
écart type <sup>2</sup>	13,39	2,98	6,19	1,37
écart type <sup>3</sup>	13,56	3,03	6,20	1,37
beta1 (min,max)	4,00 (2,15;5,69)	4,00 (3,52;4,49)	4,00 (3,11;4,79)	4,00 (3,81;4,22)
écart type <sup>1</sup>	0,53	0,13	0,24	0,06
écart type <sup>2</sup>	0,51	0,13	0,24	0,06
écart type <sup>3</sup>	0,52	0,13	0,24	0,06
beta2 (min,max)	-3,00 (-4,41;-1,26)	-3,00 (-3,61;-2,32)	-3,00 (-3,62;-2,35)	-3,00 (-3,25;-2,76)
écart type <sup>1</sup>	0,45	0,17	0,20	0,08
écart type <sup>2</sup>	0,43	0,16	0,20	0,08
écart type <sup>3</sup>	0,43	0,17	0,20	0,08
beta3 (min,max)	-5,00 (-5,84;-4,23)	-5,00 (-5,25;-4,75)	-5,00 (-5,37;-4,62)	-5,00 (-5,12;-4,88)
écart type <sup>1</sup>	0,24	0,07	0,11	0,03
écart type <sup>2</sup>	0,24	0,07	0,11	0,03
écart type <sup>3</sup>	0,24	0,07	0,11	0,03
beta4 (min,max)	9,00 (7,62;10,30)	9,00 (8,41;9,52)	9,00 (8,41;9,65)	9,00 (8,68;9,24)
écart type <sup>1</sup>	0,40	0,16	0,18	0,07
écart type <sup>2</sup>	0,39	0,16	0,18	0,07
écart type <sup>3</sup>	0,39	0,16	0,18	0,07

<sup>1</sup> écart type obtenu lors des simulations<sup>2</sup> écart type obtenu avec l'estimateur de Liang et Zeger (1986)<sup>3</sup> écart type obtenu avec l'estimateur de Prentice (1988)

TABLEAU 3.2

Corrélation des observations: Uniforme

Corrélation de départ: Uniforme

Niveau de corrélation: 0,45

Type de données: Normale

	# de sujets:=20		# de sujets:=100	
	# de répétitions par sujet:=5	# de répétitions par sujet:=10	# de répétitions par sujet:=5	# de répétitions par sujet:=10
beta0 (min,max)	2,03 (-33,76;40,59)	1,98 (-8,61;10,25)	1,97 (-14,09;19,75)	2,01 (-1,88;5,60)
biais relatif	0,02	-0,01	-0,02	0,01
écart type <sup>1</sup>	11,22	2,50	5,06	1,16
écart type <sup>2</sup>	10,79	2,48	5,00	1,14
écart type <sup>3</sup>	10,92	2,52	5,00	1,14
beta1 (min,max)	4,00 (2,52;5,34)	4,00 (3,60;4,39)	4,00 (3,28;4,62)	4,00 (3,85;4,17)
écart type <sup>1</sup>	0,43	0,10	0,19	0,05
écart type <sup>2</sup>	0,41	0,10	0,19	0,05
écart type <sup>3</sup>	0,42	0,10	0,19	0,05
beta2 (min,max)	-3,00 (-4,14;-1,67)	-3,00 (-3,47;-2,45)	-3,00 (-3,49;-2,47)	-3,00 (-3,22;-2,81)
écart type <sup>1</sup>	0,36	0,14	0,16	0,06
écart type <sup>2</sup>	0,34	0,13	0,16	0,06
écart type <sup>3</sup>	0,35	0,14	0,16	0,06
beta3 (min,max)	-5,00 (-5,69;-4,39)	-5,00 (-5,21;-4,79)	-5,00 (-5,29;-4,70)	-5,00 (-5,10;-4,91)
écart type <sup>1</sup>	0,20	0,60	0,09	0,03
écart type <sup>2</sup>	0,19	0,06	0,09	0,03
écart type <sup>3</sup>	0,19	0,06	0,09	0,03
beta4 (min,max)	9,00 (7,92;10,07)	9,00 (8,52;9,42)	9,00 (8,52;9,53)	9,00 (8,74;9,21)
écart type <sup>1</sup>	0,32	0,13	0,15	0,06
écart type <sup>2</sup>	0,31	0,13	0,14	0,06
écart type <sup>3</sup>	0,32	0,13	0,14	0,06

<sup>1</sup> écart type obtenu lors des simulations

<sup>2</sup> écart type obtenu avec l'estimateur de Liang et Zeger (1986)

<sup>3</sup> écart type obtenu avec l'estimateur de Prentice (1988)

TABLEAU 3.3

Corrélation des observations: Uniforme

Corrélation de départ: Uniforme

Niveau de corrélation: 0,15

Type de données: Binaire

	# de sujets:=20		# de sujets:=100	
	# de répétitions par sujet:=5	# de répétitions par sujet:=10	# de répétitions par sujet:=5	# de répétitions par sujet:=10
beta0 (min,max)	0,13 (-11,14;49,21)	0,01 (-1,39;1,43)	0,01 (-1,48;1,56)	-0,00 (-0,39;0,41)
biais relatif	0,13	0,01	0,01	0,00
écart type <sup>1</sup>	3,94	0,42	0,49	0,11
écart type <sup>2</sup>	2,24	0,45	0,89	0,20
écart type <sup>3</sup>	44,04	0,46	0,89	0,20
beta1 (min,max)	0,08 (-24,74;5,35)	0,14 (-0,37;0,74)	0,13 (-0,62;0,79)	0,14 (-0,00;0,32)
écart type <sup>1</sup>	1,91	0,17	0,24	0,05
écart type <sup>2</sup>	1,13	0,19	0,45	0,09
écart type <sup>3</sup>	22,30	0,20	0,45	0,09
beta2 (min,max)	-0,11 (-12,31;3,87)	-0,11 (-0,37;0,22)	-0,11 (-0,54;0,34)	-0,10 (-0,18;-0,02)
écart type <sup>1</sup>	1,25	0,1	0,15	0,03
écart type <sup>2</sup>	0,80	0,11	0,32	0,05
écart type <sup>3</sup>	10,84	0,11	0,32	0,05
beta3 (min,max)	-0,93 (-12,22;1,03)	-0,88 (-1,65;-0,20)	-0,85 (-1,33;-0,40)	-0,85 (-1,11;-0,58)
écart type <sup>1</sup>	0,8	0,27	0,14	0,08
écart type <sup>2</sup>	0,59	0,30	0,24	0,13
écart type <sup>3</sup>	10,63	0,31	0,24	0,13
beta4 (min,max)	1,10 (-9,06;34,22)	1,12 (-1,30;3,67)	1,11 (-0,11;2,26)	1,09 (0,42;1,84)
écart type <sup>1</sup>	3,38	0,71	0,41	0,23
écart type <sup>2</sup>	2,17	0,84	0,86	0,37
écart type <sup>3</sup>	29,29	0,86	0,86	0,37

<sup>1</sup> écart type obtenu lors des simulations

<sup>2</sup> écart type obtenu avec l'estimateur de Liang et Zeger (1986)

<sup>3</sup> écart type obtenu avec l'estimateur de Prentice (1988)

TABLEAU 3.4

Corrélation des observations: Uniforme

Corrélation de départ: Uniforme

Niveau de corrélation: 0,30

Type de données: Binaire

	# de sujets:=20		# de sujets:=100	
	# de répétitions par sujet:=5	# de répétitions par sujet:=10	# de répétitions par sujet:=5	# de répétitions par sujet:=10
beta0 (min,max)	0,03 (-9,30;15,30)	0,01 (-1,96;1,97)	0,00 (-1,54;2,03)	-0,00 (-0,39;0,41)
biais relatif	0,03	0,01	0,00	0,00
écart type <sup>1</sup>	2,94	0,57	0,57	0,11
écart type <sup>2</sup>	2,04	0,46	0,81	0,20
écart type <sup>3</sup>	9,25	0,46	0,81	0,20
beta1 (min,max)	0,15 (-3,50;4,53)	0,15 (-0,41;0,69)	0,13 (-0,83;0,97)	0,14 (-0,00;0,32)
écart type <sup>1</sup>	1,31	0,18	0,28	0,05
écart type <sup>2</sup>	1,02	0,18	0,40	0,09
écart type <sup>3</sup>	2,24	0,19	0,41	0,09
beta2 (min,max)	-0,08 (-2,53;3,27)	-0,11 (-0,41;0,19)	-0,10 (-0,67;0,46)	-0,10 (-0,18;-0,02)
écart type <sup>1</sup>	0,94	0,1	0,18	0,03
écart type <sup>2</sup>	0,74	0,10	0,29	0,05
écart type <sup>3</sup>	1,07	0,10	0,29	0,05
beta3 (min,max)	-0,94 (-10,43;0,95)	-0,91 (-1,79;-0,14)	-0,85 (-1,46;-0,34)	-0,85 (-1,11;-0,58)
écart type <sup>1</sup>	0,65	0,27	0,18	0,08
écart type <sup>2</sup>	0,54	0,29	0,22	0,13
écart type <sup>3</sup>	4,89	0,29	0,22	0,13
beta4 (min,max)	1,04 (-26,98;8,38)	1,16 (-1,59;4,21)	1,11 (-0,37;0,266)	1,09 (0,42;1,84)
écart type <sup>1</sup>	2,75	0,78	0,5	0,23
écart type <sup>2</sup>	2,00	0,79	0,78	0,37
écart type <sup>3</sup>	15,57	0,81	0,78	0,37

<sup>1</sup> écart type obtenu lors des simulations

<sup>2</sup> écart type obtenu avec l'estimateur de Liang et Zeger (1986)

<sup>3</sup> écart type obtenu avec l'estimateur de Prentice (1988)



TABLEAU 4.1

Corrélation des observations: Uniforme

Corrélation de départ: AR(1)

Niveau de corrélation: 0,15

Type de données: Normale

	# de sujets:=20		# de sujets:=100	
	# de répétitions par sujet:=5	# de répétitions par sujet:=10	# de répétitions par sujet:=5	# de répétitions par sujet:=10
beta0 (min,max)	2,04 (-42,09;52,20)	1,98 (-11,50;11,97)	1,96 (-18,36;24,13)	2,02 (-2,79;6,41)
biais relatif	0,02	-0,01	-0,02	0,01
écart type <sup>1</sup>	13,93	3,03	6,29	1,40
écart type <sup>2</sup>	13,39	2,99	6,19	1,37
écart type <sup>3</sup>	13,27	3,24	6,07	1,47
beta1 (min,max)	4,00 (2,15;5,69)	4,00 (3,52;4,49)	4,00 (3,11;4,79)	4,00 (3,80;4,22)
écart type <sup>1</sup>	0,53	0,13	0,24	0,06
écart type <sup>2</sup>	0,51	0,13	0,24	0,06
écart type <sup>3</sup>	0,51	0,14	0,23	0,06
beta2 (min,max)	-3,00 (-4,41;-1,26)	-3,00 (-3,61;-2,33)	-3,00 (-3,62;-2,35)	-3,00 (-3,26;-2,76)
écart type <sup>1</sup>	0,45	0,17	0,20	0,08
écart type <sup>2</sup>	0,43	0,16	0,20	0,08
écart type <sup>3</sup>	0,43	0,17	0,20	0,08
beta3 (min,max)	-5,00 (-5,84;-4,22)	-5,00 (-5,26;-4,75)	-5,00 (-5,37;-4,62)	-5,00 (-5,12;-4,89)
écart type <sup>1</sup>	0,24	0,07	0,11	0,03
écart type <sup>2</sup>	0,24	0,07	0,11	0,03
écart type <sup>3</sup>	0,24	0,08	0,11	0,04
beta4 (min,max)	9,00 (7,63;10,31)	9,00 (8,39;9,54)	9,00 (8,41;9,65)	9,00 (8,68;9,24)
écart type <sup>1</sup>	0,40	0,16	0,18	0,07
écart type <sup>2</sup>	0,39	0,16	0,18	0,07
écart type <sup>3</sup>	0,39	0,16	0,18	0,07

<sup>1</sup> écart type obtenu lors des simulations

<sup>2</sup> écart type obtenu avec l'estimateur de Liang et Zeger (1986)

<sup>3</sup> écart type obtenu avec l'estimateur de Prentice (1988)

TABLEAU 4.2

Corrélation des observations: Uniforme

Corrélation de départ: AR(1)

Niveau de corrélation: 0,45

Type de données: Normale

	# de sujets:=20		# de sujets:=100	
	# de répétitions par sujet:=5	# de répétitions par sujet:=10	# de répétitions par sujet:=5	# de répétitions par sujet:=10
beta0 (min,max)	2,03 (-33,76;40,59)	1,98 (-9,54;11,28)	1,97 (-14,09;19,75)	2,01 (-2,13;6,14)
biais relatif	0,02	-0,01	-0,02	0,01
écart type <sup>1</sup>	11,22	2,61	5,06	1,21
écart type <sup>2</sup>	10,79	2,59	4,99	1,19
écart type <sup>3</sup>	10,20	2,90	4,66	1,32
beta1 (min,max)	4,00 (2,52;5,34)	4,00 (3,55;4,41)	4,00 (3,28;4,62)	4,00 (3,82;4,18)
écart type <sup>1</sup>	0,43	0,11	0,19	0,05
écart type <sup>2</sup>	0,41	0,11	0,19	0,05
écart type <sup>3</sup>	0,40	0,13	0,18	0,06
beta2 (min,max)	-3,00 (-4,14;-1,68)	-3,00 (-3,51;-2,48)	-3,00 (-3,49;-2,47)	-3,00 (-3,21;-2,80)
écart type <sup>1</sup>	0,36	0,14	0,16	0,06
écart type <sup>2</sup>	0,34	0,13	0,16	0,06
écart type <sup>3</sup>	0,33	0,14	0,15	0,06
beta3 (min,max)	-5,00 (-5,69;-4,39)	-5,00 (-5,23;-4,77)	-5,00 (-5,29;-4,70)	-5,00 (-5,10;-4,90)
écart type <sup>1</sup>	0,20	0,06	0,09	0,03
écart type <sup>2</sup>	0,19	0,06	0,09	0,03
écart type <sup>3</sup>	0,20	0,08	0,09	0,04
beta4 (min,max)	9,00 (7,92;10,07)	9,00 (8,48;9,42)	9,00 (8,52;9,53)	9,00 (8,75;9,21)
écart type <sup>1</sup>	0,32	0,13	0,15	0,06
écart type <sup>2</sup>	0,31	0,13	0,14	0,06
écart type <sup>3</sup>	0,30	0,14	0,14	0,06

<sup>1</sup> écart type obtenu lors des simulations

<sup>2</sup> écart type obtenu avec l'estimateur de Liang et Zeger (1986)

<sup>3</sup> écart type obtenu avec l'estimateur de Prentice (1988)

TABLEAU 4.3

Corrélation des observations: Uniforme

Corrélation de départ: AR(1)

Niveau de corrélation: 0,15

Type de données: Binaire

	# de sujets:=20		# de sujets:=100	
	# de répétitions	# de répétitions	# de répétitions	# de répétitions
	par sujet:=5	par sujet:=10	par sujet:=5	par sujet:=10
beta0 (min,max)	0,04 (-11,14;9,20)	0,01 (-1,47;1,54)	0,01 (-1,48;1,56)	-0,00 (-0,40;0,40)
biais relatif	0,04	0,01	0,01	0,00
écart type <sup>1</sup>	3,32	0,47	0,49	0,11
écart type <sup>2</sup>	2,18	0,45	0,89	0,20
écart type <sup>3</sup>	2,12	0,46	0,86	0,20
beta1 (min,max)	0,13 (-4,48;5,35)	0,14 (-0,36;0,73)	0,13 (-0,62;0,79)	0,14 (-0,01;0,32)
écart type <sup>1</sup>	1,58	0,18	0,24	0,05
écart type <sup>2</sup>	1,10	0,20	0,45	0,09
écart type <sup>3</sup>	1,08	0,22	0,44	0,10
beta2 (min,max)	-0,08 (-3,05;3,87)	-0,10 (-0,38;0,23)	-0,11 (-0,54;0,34)	-0,10 (-0,19;-0,02)
écart type <sup>1</sup>	1,14	0,1	0,15	0,02
écart type <sup>2</sup>	0,79	0,11	0,32	0,05
écart type <sup>3</sup>	0,80	0,12	0,32	0,05
beta3 (min,max)	-0,91 (-2,85;1,03)	-0,88 (-1,65;-0,16)	-0,85 (-1,33;-0,40)	-0,85 (-1,12;-0,58)
écart type <sup>1</sup>	0,62	0,27	0,14	0,08
écart type <sup>2</sup>		0,30	0,24	0,13
écart type <sup>3</sup>	0,56	0,31	0,24	0,13
beta4 (min,max)	1,04 (-9,06;9,66)	1,11 (-1,22;3,50)	1,11 (-0,11;2,26)	1,09 (0,48;1,87)
écart type <sup>1</sup>	3,07	0,68	0,41	0,23
écart type <sup>2</sup>	2,14	0,85	0,86	0,38
écart type <sup>3</sup>	2,28	0,93	0,93	0,41

<sup>1</sup> écart type obtenu lors des simulations

<sup>2</sup> écart type obtenu avec l'estimateur de Liang et Zeger (1986)

<sup>3</sup> écart type obtenu avec l'estimateur de Prentice (1988)

TABLEAU 4.4

Corrélation des observations: Uniforme

Corrélation de départ: AR(1)

Niveau de corrélation: 0,30

Type de données: Binaire

	# de sujets:=20		# de sujets:=100	
	# de répétitions par sujet:=5	# de répétitions par sujet:=10	# de répétitions par sujet:=5	# de répétitions par sujet:=10
beta0 (min,max)	0,03 (-9,30;15,30)	0,00 (-2,11;1,98)	0,00 (-1,54;2,03)	-0,00 (-0,41;0,45)
biais relatif	0,03	0,00	0,00	0,00
écart type <sup>1</sup>	2,94	0,65	0,57	0,13
écart type <sup>2</sup>	2,03	0,47	0,81	0,21
écart type <sup>3</sup>	9,12	0,45	0,76	0,19
beta1 (min,max)	0,15 (-3,50;4,53)	0,15 (-0,44;0,71)	0,13 (-0,83;0,97)	0,14 (-0,05;0,31)
écart type <sup>1</sup>	1,31	0,19	0,28	0,05
écart type <sup>2</sup>	1,02	0,19	0,40	0,08
écart type <sup>3</sup>	2,22	0,22	0,39	0,09
beta2 (min,max)	-0,08 (-2,53;3,27)	-0,11 (-0,41;0,18)	-0,10 (-0,67;0,46)	-0,10 (-0,20;-0,01)
écart type <sup>1</sup>	0,94	0,1	0,18	0,03
écart type <sup>2</sup>	0,74	0,11	0,29	0,05
écart type <sup>3</sup>	1,10	0,13	0,30	0,05
beta3 (min,max)	-0,94 (-10,43;0,95)	-0,91 (-1,79;-0,13)	-0,85 (-1,46;-0,34)	-0,85 (-1,17;-0,53)
écart type <sup>1</sup>	0,65	0,28	0,18	0,09
écart type <sup>2</sup>	0,53	0,29	0,22	0,12
écart type <sup>3</sup>	4,87	0,28	0,22	0,12
beta4 (min,max)	1,04 (-26,98;8,38)	1,15 (-1,77;4,22)	1,11 (-0,37;2,66)	1,10 (0,25;1,89)
écart type <sup>1</sup>	2,74	0,77	0,5	0,25
écart type <sup>2</sup>	2,00	0,83	0,78	0,36
écart type <sup>3</sup>	15,89	0,93	0,93	0,40

<sup>1</sup> écart type obtenu lors des simulations

<sup>2</sup> écart type obtenu avec l'estimateur de Liang et Zeger (1986)

<sup>3</sup> écart type obtenu avec l'estimateur de Prentice (1988)

TABLEAU 5.1

Corrélation des observations: Uniforme

Corrélation de départ: Indépendante

Niveau de corrélation: 0,15

Type de données: Normale

	# de sujets:=20		# de sujets:=100	
	# de répétitions par sujet:=5	# de répétitions par sujet:=10	# de répétitions par sujet:=5	# de répétitions par sujet:=10
beta0 (min,max)	2,04 (-42,09;52,20)	1,98 (-11,19;12,16)	1,96 (-18,36;24,13)	2,02 (-2,66;6,22)
biais relatif	0,02	-0,01	-0,02	0,01
écart type <sup>1</sup>	13,93	3,01	6,29	1,39
écart type <sup>2</sup>	13,39	2,98	6,19	1,37
écart type <sup>3</sup>	14,69	3,25	6,72	1,47
beta1 (min,max)	4,00 (2,15;5,69)	4,00 (3,52;4,49)	4,00 (3,11;4,79)	4,00 (3,81;4,22)
écart type <sup>1</sup>	0,53	0,13	0,24	0,06
écart type <sup>2</sup>	0,51	0,13	0,24	0,06
écart type <sup>3</sup>	0,56	0,14	0,26	0,06
beta2 (min,max)	-3,00 (-4,41;-1,26)	-3,00 (-3,61;-2,32)	-3,00 (-3,62;-2,35)	-3,00 (-3,26;-2,76)
écart type <sup>1</sup>	0,45	0,17	0,20	0,08
écart type <sup>2</sup>	0,43	0,16	0,20	0,08
écart type <sup>3</sup>	0,47	0,18	0,22	0,08
beta3 (min,max)	-5,00 9-5,84;-4,23)	-5,00 (-5,25;-4,75)	-5,00 (-5,37;-4,62)	-5,00 (-5,12;-4,89)
écart type <sup>1</sup>	0,24	0,07	0,11	0,03
écart type <sup>2</sup>	0,24	0,07	0,11	0,03
écart type <sup>3</sup>	0,26	0,08	0,12	0,04
beta4 (min,max)	9,00 (7,63;10,31)	9,00 (8,41;9,52)	9,00 (8,41;9,65)	9,00 (8,68;9,24)
écart type <sup>1</sup>	0,40	0,16	0,18	0,07
écart type <sup>2</sup>	0,39	0,16	0,18	0,07
écart type <sup>3</sup>	0,43	0,17	0,19	0,08

<sup>1</sup> écart type obtenu lors des simulations

<sup>2</sup> écart type obtenu avec l'estimateur de Liang et Zeger (1986)

<sup>3</sup> écart type obtenu avec l'estimateur de Prentice (1988)

TABLEAU 5.2

Corrélation des observations: Uniforme

Corrélation de départ: Indépendante

Niveau de corrélation: 0,45

Type de données: Normale

	# de sujets:=20		# de sujets:=100	
	# de répétitions par sujet:=5	# de répétitions par sujet:=10	# de répétitions par sujet:=5	# de répétitions par sujet:=10
beta0 (min,max)	2,03 (-33,76;40,59)	1,98 (-8,61;10,25)	1,97 (-14,09;19,75)	2,01 (-1,88;5,60)
biais relatif	0,02	-0,01	-0,02	0,01
écart type <sup>1</sup>	11,22	2,50	5,06	1,16
écart type <sup>2</sup>	10,79	2,48	4,99	1,14
écart type <sup>3</sup>	14,66	3,23	6,72	1,47
beta1 (min,max)	4,00 (2,52;5,34)	4,00 (3,61;4,40)	4,00 (3,28;4,62)	4,00 (3,85;4,17)
écart type <sup>1</sup>	0,43	0,10	0,19	0,05
écart type <sup>2</sup>	0,41	0,10	0,19	0,05
écart type <sup>3</sup>	0,56	0,14	0,26	0,06
beta2 (min,max)	-3,00 (-4,14;-1,68)	-3,00 (-3,47;-2,45)	-3,00 (-3,49;-2,47)	-3,00 (-3,22;-2,81)
écart type <sup>1</sup>	0,36	0,14	0,16	0,06
écart type <sup>2</sup>	0,34	0,13	0,16	0,06
écart type <sup>3</sup>	0,47	0,18	0,22	0,08
beta3 (min,max)	-5,00 (-5,69;-4,39)	-5,00 (-5,21;-4,79)	-5,00 (-5,29;-4,70)	-5,00 (-5,10;-4,91)
écart type <sup>1</sup>	0,20	0,06	0,09	0,03
écart type <sup>2</sup>	0,19	0,06	0,09	0,03
écart type <sup>3</sup>	0,26	0,08	0,12	0,04
beta4 (min,max)	9,00 (7,92;10,07)	9,00 (8,52;9,42)	9,00 (8,52;9,53)	9,00 (8,74;9,21)
écart type <sup>1</sup>	0,32	0,13	0,15	0,06
écart type <sup>2</sup>	0,31	0,13	0,14	0,06
écart type <sup>3</sup>	0,42	0,17	0,19	0,08

<sup>1</sup> écart type obtenu lors des simulations

<sup>2</sup> écart type obtenu avec l'estimateur de Liang et Zeger (1986)

<sup>3</sup> écart type obtenu avec l'estimateur de Prentice (1988)

TABLEAU 5.3

Corrélation des observations: Uniforme

Corrélation de départ: Indépendante

Niveau de corrélation: 0,15

Type de données: Binaire

	# de sujets:=20		# de sujets:=100	
	# de répétitions par sujet:=5	# de répétitions par sujet:=10	# de répétitions par sujet:=5	# de répétitions par sujet:=10
beta0 (min,max)	0,09 (-11,14;49,20)	0,01 (-1,38;1,43)	0,01 (-1,48;1,56)	-0,00 (-0,39;0,41)
biais relatif	0,09	0,01	0,01	0,00
écart type <sup>1</sup>	3,66	0,42	0,49	0,11
écart type <sup>2</sup>	2,23	0,45	0,89	0,20
écart type <sup>3</sup>	23,29	0,46	0,96	0,20
beta1 (min,max)	0,10 (-24,74;5,35)	0,14 (-0,37;0,74)	0,13 (-0,61;0,79)	0,14 (-0,00;0,32)
écart type <sup>1</sup>	1,76	0,17	0,24	0,05
écart type <sup>2</sup>	1,12	0,19	0,45	0,09
écart type <sup>3</sup>	11,79	0,21	0,48	0,09
beta2 (min,max)	-0,10 (-12,31;3,87)	-0,10 (-0,37;0,22)	-0,11 (-0,54;0,34)	-0,10 (-0,18;-0,02)
écart type <sup>1</sup>	1,2	0,1	0,15	0,03
écart type <sup>2</sup>	0,80	0,11	0,32	0,05
écart type <sup>3</sup>	5,88	0,12	0,35	0,05
beta3 (min,max)	-0,92 (-12,22;1,03)	-0,88 (-1,65;-0,20)	-0,85 (-1,33;-0,40)	-0,85 (-1,11;-0,58)
écart type <sup>1</sup>	0,72	0,27	0,14	0,08
écart type <sup>2</sup>	0,58	0,30	0,24	0,13
écart type <sup>3</sup>	5,65	0,33	0,27	0,15
beta4 (min,max)	1,07 (-9,06;34,22)	1,12 (-1,30;3,66)	1,11 (-0,11;2,26)	1,10 (0,42;1,84)
écart type <sup>1</sup>	3,24	0,71	0,41	0,23
écart type <sup>2</sup>	2,16	0,84	0,86	0,37
écart type <sup>3</sup>	15,88	0,92	0,93	0,40

<sup>1</sup> écart type obtenu lors des simulations

<sup>2</sup> écart type obtenu avec l'estimateur de Liang et Zeger (1986)

<sup>3</sup> écart type obtenu avec l'estimateur de Prentice (1988)

TABLEAU 5.4

Corrélation des observations: Uniforme

Corrélation de départ: Indépendante

Niveau de corrélation: 0,30

Type de données: Binaire

	# de sujets:=20		# de sujets:=100	
	# de répétitions par sujet:=5	# de répétitions par sujet:=10	# de répétitions par sujet:=5	# de répétitions par sujet:=10
beta0 (min,max)	0,03 (-9,30;15,30)	0,01 (-1,95;2,42)	0,00 (-1,54;2,03)	-0,00 (-0,41;0,42)
biais relatif	0,03	0,01	0,00	0,00
écart type <sup>1</sup>	2,94	0,57	0,57	0,12
écart type <sup>2</sup>	2,04	0,46	0,81	0,20
écart type <sup>3</sup>	9,57	0,47	0,96	0,20
beta1 (min,max)	0,15 (-3,50;4,52)	0,15 (-0,62;0,61)	0,13 (-0,83;0,97)	0,14 (-0,03;0,28)
écart type <sup>1</sup>	1,3	0,18	0,28	0,05
écart type <sup>2</sup>	1,02	0,18	0,40	0,08
écart type <sup>3</sup>	2,41	0,22	0,48	0,09
beta2 (min,max)	-0,08 (2,53;3,27)	-0,11 (-0,40;0,17)	-0,10 (-0,67;0,46)	-0,10 (-0,19;-0,01)
écart type <sup>1</sup>	0,93	0,1	0,18	0,03
écart type <sup>2</sup>	0,74	0,10	0,29	0,04
écart type <sup>3</sup>	1,19	0,12	0,35	0,05
beta3 (min,max)	-0,94 (-10,42;0,95)	-0,92 (-1,86;-0,23)	-0,85 (-1,46;-0,34)	-0,85 (-1,17;-0,52)
écart type <sup>1</sup>	0,65	0,28	0,18	0,09
écart type <sup>2</sup>	0,54	0,29	0,22	0,12
écart type <sup>3</sup>	4,98	0,34	0,27	0,15
beta4 (min,max)	1,04 (-26,98;8,37)	1,13 (-1,02;3,74)	1,11 (-0,36;2,66)	1,10 (0,39;1,85)
écart type <sup>1</sup>	2,75	0,75	0,5	0,23
écart type <sup>2</sup>	2,01	0,79	0,78	0,34
écart type <sup>3</sup>	15,91	0,94	0,93	0,40

<sup>1</sup> écart type obtenu lors des simulations

<sup>2</sup> écart type obtenu avec l'estimateur de Liang et Zeger (1986)

<sup>3</sup> écart type obtenu avec l'estimateur de Prentice (1988)



TABLEAU 6.1

Corrélation des observations: AR(1)

Corrélation de départ: Uniforme

Niveau de corrélation: 0,15

Type de données: Normale

	# de sujets:=20		# de sujets:=100	
	# de répétitions par sujet:=5	# de répétitions par sujet:=10	# de répétitions par sujet:=5	# de répétitions par sujet:=10
beta0 (min,max)	2,03 (-40,12;46,97)	1,98 (-12,08;12,56)	1,96 (-17,21;23,38)	2,02 (-3,04;6,70)
biais relatif	0,01	-0,01	-0,02	0,01
écart type <sup>1</sup>	13,59	3,29	6,15	1,52
écart type <sup>2</sup>	13,07	3,26	6,05	1,49
écart type <sup>3</sup>	14,25	3,21	6,52	1,45
beta1 (min,max)	4,00 (2,27;5,64)	4,00 (3,49;4,53)	4,00 (3,14;4,75)	4,00 (3,78;4,24)
écart type <sup>1</sup>	0,52	0,14	0,24	0,07
écart type <sup>2</sup>	0,50	0,14	0,23	0,07
écart type <sup>3</sup>	0,55	0,14	0,25	0,06
beta2 (min,max)	-3,00 (-4,35;-1,41)	-3,00 (-3,65;-2,30)	-3,00 (-3,59;-2,34)	-3,00 (-3,27;-2,76)
écart type <sup>1</sup>	0,44	0,18	0,20	0,08
écart type <sup>2</sup>	0,42	0,17	0,20	0,08
écart type <sup>3</sup>	0,46	0,18	0,21	0,08
beta3 (min,max)	-5,00 (-5,90;-4,23)	-5,00 (-5,28;-4,72)	-5,00 (-5,38;-4,62)	-5,00 (-5,14;-4,87)
écart type <sup>1</sup>	0,25	0,08	0,11	0,04
écart type <sup>2</sup>	0,24	0,08	0,11	0,04
écart type <sup>3</sup>	0,25	0,08	0,12	0,04
beta4 (min,max)	9,00 (7,66;10,26)	9,00 (8,39;9,53)	9,00 (8,42;9,63)	9,00 (8,68;9,25)
écart type <sup>1</sup>	0,39	0,16	0,18	0,08
écart type <sup>2</sup>	0,38	0,16	0,17	0,07
écart type <sup>3</sup>	0,41	0,17	0,19	0,08

<sup>1</sup> écart type obtenu lors des simulations

<sup>2</sup> écart type obtenu avec l'estimateur de Liang et Zeger (1986)

<sup>3</sup> écart type obtenu avec l'estimateur de Prentice (1988)

TABLEAU 6.2

Corrélation des observations: AR(1)

Corrélation de départ: Uniforme

Niveau de corrélation: 0,45

Type de données: Normale

	# de sujets:=20		# de sujets:=100	
	# de répétitions par sujet:=5	# de répétitions par sujet:=10	# de répétitions par sujet:=5	# de répétitions par sujet:=10
beta0 (min,max)	2,01 (-27,52;31,76)	1,98 (-10,28;12,97)	1,96 (-13,09;17,68)	2,01 (-3,02;7,56)
biais relatif	0,01	-0,01	-0,02	0,01
écart type <sup>1</sup>	9,86	3,41	4,50	1,56
écart type <sup>2</sup>	9,58	3,37	4,44	1,55
écart type <sup>3</sup>	13,30	3,12	6,10	1,41
beta1 (min,max)	4,00 (2,81;5,24)	4,00 (3,48;4,51)	4,00 (3,40;4,60)	4,00 (3,75;4,22)
écart type <sup>1</sup>	0,38	0,15	0,17	0,07
écart type <sup>2</sup>	0,37	0,15	0,17	0,07
écart type <sup>3</sup>	0,51	0,13	0,23	0,06
beta2 (min,max)	-3,00 (-3,97;-2,02)	-3,00 (-3,58;-2,37)	-3,00 (-3,45;-2,46)	-3,00 (-3,25;-2,78)
écart type <sup>1</sup>	0,33	0,16	0,15	0,07
écart type <sup>2</sup>	0,32	0,15	0,15	0,07
écart type <sup>3</sup>	0,43	0,17	0,20	0,08
beta3 (min,max)	-5,00 (-5,85;-4,37)	-5,00 (-5,31;-4,68)	-5,00 (-5,30;-4,68)	-5,00 (-5,15;-4,85)
écart type <sup>1</sup>	0,21	0,09	0,09	0,04
écart type <sup>2</sup>	0,20	0,09	0,09	0,04
écart type <sup>3</sup>	0,23	0,08	0,11	0,03
beta4 (min,max)	9,00 (7,87;9,83)	9,00 (8,44;9,48)	9,00 (8,59;9,44)	9,00 (8,73;9,21)
écart type <sup>1</sup>	0,28	0,14	0,13	0,06
écart type <sup>2</sup>	0,27	0,14	0,12	0,06
écart type <sup>3</sup>	0,39	0,17	0,18	0,08

<sup>1</sup> écart type obtenu lors des simulations

<sup>2</sup> écart type obtenu avec l'estimateur de Liang et Zeger (1986)

<sup>3</sup> écart type obtenu avec l'estimateur de Prentice (1988)

TABLEAU 6.3

Corrélation des observations: AR(1)

Corrélation de départ: Uniforme

Niveau de corrélation: 0,15

Type de données: Binaire

	# de sujets:=20		# de sujets:=100	
	# de répétitions par sujet:=5	# de répétitions par sujet:=10	# de répétitions par sujet:=5	# de répétitions par sujet:=10
beta0 (min,max)	0,06 (-9,55;8,99)	0,01 (-1,11;1,05)	0,00 (-1,38;1,42)	0,00 (-0,35;0,38)
biais relatif	0,06	0,01	0,00	0,00
écart type <sup>1</sup>	3,09	0,31	0,41	0,1
écart type <sup>2</sup>	2,07	0,46	0,84	0,20
écart type <sup>3</sup>	2,29	0,45	0,94	0,20
beta1 (min,max)	0,12 (-4,81;5,18)	0,14 (-0,35;0,67)	0,14 (-0,48;0,85)	0,14 (-0,04;0,28)
écart type <sup>1</sup>	1,57	0,19	0,2	0,05
écart type <sup>2</sup>	1,06	0,22	0,43	0,10
écart type <sup>3</sup>	1,15	0,21	0,47	0,09
beta2 (min,max)	-0,09 (-3,30;3,76)	-0,10 (-0,40;0,24)	-0,10 (-0,49;0,37)	-0,10 (-0,18;-0,04)
écart type <sup>1</sup>	1,2	0,11	0,13	0,02
écart type <sup>2</sup>	0,79	0,12	0,32	0,05
écart type <sup>3</sup>	0,82	0,12	0,34	0,05
beta3 (min,max)	-0,91 (-2,76;0,54)	-0,87 (-1,63;-0,35)	-0,85 (-1,22;-0,45)	-0,85 (-1,14;-0,65)
écart type <sup>1</sup>	0,53	0,23	0,12	0,07
écart type <sup>2</sup>	0,56	0,29	0,24	0,13
écart type <sup>3</sup>	0,60	0,33	0,26	0,14
beta4 (min,max)	1,05 (-10,44;11,33)	1,11 (-1,43;3,58)	1,09 (0,05;2,47)	1,10 (-0,43;1,87)
écart type <sup>1</sup>	3,5	0,8	0,35	0,23
écart type <sup>2</sup>	2,28	0,93	0,93	0,42
écart type <sup>3</sup>	2,22	0,90	0,90	0,40

<sup>1</sup> écart type obtenu lors des simulations

<sup>2</sup> écart type obtenu avec l'estimateur de Liang et Zeger (1986)

<sup>3</sup> écart type obtenu avec l'estimateur de Prentice (1988)

TABLEAU 6.4

Corrélation des observations: AR(1)

Corrélation de départ: Uniforme

Niveau de corrélation: 0,30

Type de données: Binaire

	# de sujets:=20		# de sujets:=100	
	# de répétitions	# de répétitions	# de répétitions	# de répétitions
	par sujet:=5	par sujet:=10	par sujet:=5	par sujet:=10
beta0 (min,max)	0,02 (-8,11;45,32)	-0,01 (-1,35;1,17)	-0,01 (-1,77;1,55)	0,00 (-0,40;0,40)
biais relatif	0,02	-0,01	-0,01	0,00
écart type <sup>1</sup>	2,75	0,42	0,48	0,13
écart type <sup>2</sup>	1,84	0,47	0,70	0,21
écart type <sup>3</sup>	22,03	0,46	0,91	0,20
beta1 (min,max)	0,13 (-23,65;4,71)	0,15 (-0,57;1,00)	0,14 (-0,61;1,05)	0,14 (-0,04;0,32)
écart type <sup>1</sup>	1,43	0,25	0,24	0,06
écart type <sup>2</sup>	0,97	0,23	0,37	0,10
écart type <sup>3</sup>	11,25	0,21	0,45	0,09
beta2 (min,max)	-0,10 (-20,59;3,42)	-0,11 (-0,49;0,33)	-0,10 (-0,59;0,52)	-0,10 (-0,21;0,00)
écart type <sup>1</sup>	1,27	0,13	0,18	0,03
écart type <sup>2</sup>	0,76	0,13	0,29	0,06
écart type <sup>3</sup>	9,43	0,13	0,32	0,05
beta3 (min,max)	-0,94 (-21,66;0,37)	-0,87 (-1,59;-0,11)	-0,85 (-1,42;-0,23)	-0,85 (-1,19;-0,62)
écart type <sup>1</sup>	0,82	0,23	0,18	0,09
écart type <sup>2</sup>	0,52	0,26	0,21	0,12
écart type <sup>3</sup>	9,21	0,32	0,25	0,14
beta4 (min,max)	1,14 (-10,18;57,08)	1,14 (-1,68;4,16)	1,09 (-0,44;2,74)	1,10 (0,18;2,04)
écart type <sup>1</sup>	3,93	1,04	0,54	0,28
écart type <sup>2</sup>	2,37	0,97	0,93	0,43
écart type <sup>3</sup>	25,47	0,90	0,88	0,39

<sup>1</sup> écart type obtenu lors des simulations

<sup>2</sup> écart type obtenu avec l'estimateur de Liang et Zeger (1986)

<sup>3</sup> écart type obtenu avec l'estimateur de Prentice (1988)

TABLEAU 7.1

Corrélation des observations: AR(1)

Corrélation de départ: AR(1)

Niveau de corrélation: 0,15

Type de données: Normale

	# de sujets:=20		# de sujets:=100	
	# de répétitions par sujet:=5	# de répétitions par sujet:=10	# de répétitions par sujet:=5	# de répétitions par sujet:=10
beta0 (min,max)	2,03 (-40,12;46,97)	1,98 (-12,47;12,69)	1,96 (-17,21;23,38)	2,02 (-3,19;6,71)
biais relatif	0,02	-0,01	-0,02	0,01
écart type <sup>1</sup>	13,59	3,27	6,15	1,51
écart type <sup>2</sup>	13,07	3,23	6,05	1,48
écart type <sup>3</sup>	13,27	3,25	6,07	1,47
beta1 (min,max)	4,00 (2,27;5,64)	4,00 (3,49;4,53)	4,00 (3,14;4,75)	4,00 (3,79;4,24)
écart type <sup>1</sup>	0,52	0,14	0,24	0,07
écart type <sup>2</sup>	0,50	0,14	0,23	0,06
écart type <sup>3</sup>	0,51	0,14	0,24	0,06
beta2 (min,max)	-3,00 (-4,35;-1,41)	-3,00 (-3,64;-2,32)	-3,00 (-3,59;-2,34)	-3,00 (-3,27;-2,76)
écart type <sup>1</sup>	0,44	0,18	0,20	0,08
écart type <sup>2</sup>	0,42	0,17	0,20	0,08
écart type <sup>3</sup>	0,43	0,17	0,20	0,08
beta3 (min,max)	-5,00 (-5,90;-4,23)	-5,00 (-5,28;-4,72)	-5,00 (-5,38;-4,62)	-5,00 (-5,14;-4,87)
écart type <sup>1</sup>	0,25	0,08	0,11	0,04
écart type <sup>2</sup>	0,24	0,08	0,11	0,04
écart type <sup>3</sup>	0,24	0,08	0,11	0,04
beta4 (min,max)	9,00 (7,66;10,26)	9,00 (8,39;9,53)	9,00 (8,42;9,63)	9,00 (8,68;9,25)
écart type <sup>1</sup>	0,39	0,16	0,18	0,08
écart type <sup>2</sup>	0,38	0,16	0,17	0,07
écart type <sup>3</sup>	0,39	0,16	0,18	0,07

<sup>1</sup> écart type obtenu lors des simulations

<sup>2</sup> écart type obtenu avec l'estimateur de Liang et Zeger (1986)

<sup>3</sup> écart type obtenu avec l'estimateur de Prentice (1988)

TABLEAU 7.2

Corrélation des observations: AR(1)

Corrélation de départ: AR(1)

Niveau de corrélation: 0,45

Type de données: Normale

	# de sujets:=20		# de sujets:=100	
	# de répétitions par sujet:=5	# de répétitions par sujet:=10	# de répétitions par sujet:=5	# de répétitions par sujet:=10
beta0 (min,max)	2,00 (-27,52;31,76)	1,98 (-10,66;12,07)	1,96 (-13,09;17,68)	2,01 (-2,65;6,85)
biais relatif	0,00	-0,01	-0,02	0,01
écart type <sup>1</sup>	9,86	3,15	4,50	1,45
écart type <sup>2</sup>	9,58	3,13	4,44	1,44
écart type <sup>3</sup>	10,19	2,92	4,66	1,32
beta1 (min,max)	4,00 (2,81;5,24)	4,00 (3,50;4,46)	4,00 (3,41;4,60)	4,00 (3,76;4,22)
écart type <sup>1</sup>	0,38	0,14	0,17	0,06
écart type <sup>2</sup>	0,37	0,14	0,17	0,06
écart type <sup>3</sup>	0,40	0,13	0,18	0,06
beta2 (min,max)	-3,00 (-3,97;-2,02)	-3,00 (-3,56;-2,41)	-3,00 (-3,45;-2,46)	3,00 (-3,25;-2,78)
écart type <sup>1</sup>	0,33	0,16	0,15	0,07
écart type <sup>2</sup>	0,32	0,15	0,15	0,07
écart type <sup>3</sup>	0,33	0,14	0,15	0,06
beta3 (min,max)	-5,00 (-5,85;-4,37)	-5,00 (-5,33;-4,69)	-5,00 (-5,30;-4,68)	-5,00 (-5,13;-4,85)
écart type <sup>1</sup>	0,21	0,09	0,09	0,04
écart type <sup>2</sup>	0,20	0,09	0,09	0,04
écart type <sup>3</sup>	0,20	0,08	0,09	0,04
beta4 (min,max)	9,00 (7,87;9,83)	9,00 (8,46;9,44)	9,00 (8,59;9,44)	9,00 (8,75;9,21)
écart type <sup>1</sup>	0,28	0,14	0,13	0,06
écart type <sup>2</sup>	0,27	0,13	0,12	0,06
écart type <sup>3</sup>	0,30	0,014	0,14	0,06

<sup>1</sup> écart type obtenu lors des simulations

<sup>2</sup> écart type obtenu avec l'estimateur de Liang et Zeger (1986)

<sup>3</sup> écart type obtenu avec l'estimateur de Prentice (1988)

TABLEAU 7.3

Corrélation des observations: AR(1)

Corrélation de départ: AR(1)

Niveau de corrélation: 0,15

Type de données: Binaire

	# de sujets:=20		# de sujets:=100	
	# de répétitions par sujet:=5	# de répétitions par sujet:=10	# de répétitions par sujet:=5	# de répétitions par sujet:=10
beta0 (min,max)	0,06 (-9,54;8,99)	-0,01 (-1,29;1,07)	0,00 (-1,38;1,42)	-0,00 (-0,34;0,32)
biais relatif	0,06	-0,01	0,00	0,00
écart type <sup>1</sup>	3,09	0,33	0,41	0,11
écart type <sup>2</sup>	2,07	0,45	0,84	0,20
écart type <sup>3</sup>	2,12	0,45	0,86	0,20
beta1 (min,max)	0,12 (-4,81;5,19)	0,14 (-0,42;0,71)	0,14 (-0,48;0,85)	0,14 (-0,01;0,31)
écart type <sup>1</sup>	1,56	0,2	0,2	0,05
écart type <sup>2</sup>	1,06	0,21	0,43	0,10
écart type <sup>3</sup>	1,08	0,21	0,44	0,10
beta2 (min,max)	-0,09 (-3,30;3,76)	-0,10 (-0,46;0,25)	-0,10 (-0,49;0,37)	-0,10 (-0,18;-0,04)
écart type <sup>1</sup>	1,2	0,11	0,13	0,02
écart type <sup>2</sup>	0,79	0,12	0,32	0,05
écart type <sup>3</sup>	0,80	0,12	0,32	0,05
beta3 (min,max)	-0,91 (-2,76;0,54)	-0,87 (-1,57;-0,17)	-0,85 (-1,22;-0,44)	-0,85 (-1,09;-0,62)
écart type <sup>1</sup>	0,53	0,24	0,12	0,07
écart type <sup>2</sup>	0,56	0,30	0,24	0,13
écart type <sup>3</sup>	0,56	0,30	0,24	0,13
beta4 (min,max)	1,05 (-10,44;11,33)	1,12 (-1,21;3,27)	1,09 (0,05;2,47)	1,10 (-0,40;1,84)
écart type <sup>1</sup>	3,5	0,71	0,36	0,23
écart type <sup>2</sup>	2,27	0,91	0,93	0,41
écart type <sup>3</sup>	2,28	0,92	0,93	0,41

<sup>1</sup> écart type obtenu lors des simulations

<sup>2</sup> écart type obtenu avec l'estimateur de Liang et Zeger (1986)

<sup>3</sup> écart type obtenu avec l'estimateur de Prentice (1988)

TABLEAU 7.4

Corrélation des observations: AR(1)

Corrélation de départ: AR(1)

Niveau de corrélation: 0,30

Type de données: Binaire

	# de sujets:=20		# de sujets:=100	
	# de répétitions par sujet:=5	# de répétitions par sujet:=10	# de répétitions par sujet:=5	# de répétitions par sujet:=10
beta0 (min,max)	-0,03 (-8,11;6,07)	-0,01 (-1,07;1,43)	-0,01 (-1,77;1,54)	-0,00 (-0,42;0,41)
biais relatif	-0,03	-0,01	-0,01	0,00
écart type <sup>1</sup>	2,35	0,4	0,48	0,13
écart type <sup>2</sup>	1,80	0,44	0,70	0,20
écart type <sup>3</sup>	1,90	0,44	0,76	0,20
beta1 (min,max)	0,16 (-3,44;4,71)	0,14 (-0,56;0,99)	0,14 (-0,60;1,05)	0,14 (-0,06;0,33)
écart type <sup>1</sup>	1,21	0,24	0,24	0,06
écart type <sup>2</sup>	0,94	0,22	0,37	0,10
écart type <sup>3</sup>	0,98	0,21	0,39	0,09
beta2 (min,max)	-0,08 (-2,85;3,41)	-0,11 (-0,47;0,25)	-0,09 (-0,59;0,52)	-0,10 (-0,20;0,01)
écart type <sup>1</sup>	1,09	0,12	0,18	0,03
écart type <sup>2</sup>	0,74	0,12	0,29	0,05
écart type <sup>3</sup>	0,75	0,12	0,30	0,05
beta3 (min,max)	-0,92 (-3,26;0,37)	-0,88 (-1,67;-0,13)	-0,84 (-1,42;-0,23)	-0,85 (-1,20;-0,61)
écart type <sup>1</sup>	0,48	0,24	0,18	0,09
écart type <sup>2</sup>	0,50	0,26	0,21	0,12
écart type <sup>3</sup>	0,51	0,27	0,21	0,12
beta4 (min,max)	1,08 (-10,17;9,74)	1,15 (-1,33;4,26)	1,09 (-0,44;2,74)	1,10 (0,16;1,94)
écart type <sup>1</sup>	3,51	0,88	0,54	0,27
écart type <sup>2</sup>	2,31	0,90	0,93	0,40
écart type <sup>3</sup>	2,28	0,91	0,93	0,40

<sup>1</sup> écart type obtenu lors des simulations

<sup>2</sup> écart type obtenu avec l'estimateur de Liang et Zeger (1986)

<sup>3</sup> écart type obtenu avec l'estimateur de Prentice (1988)



TABLEAU 8.1

Corrélation des observations: AR(1)

Corrélation de départ: Indépendante

Niveau de corrélation: 0,15

Type de données: Normale

	# de sujets:=20		# de sujets:=100	
	# de répétitions par sujet:=5	# de répétitions par sujet:=10	# de répétitions par sujet:=5	# de répétitions par sujet:=10
beta0 (min,max)	2,03 (-40,12;46,97)	1,98 (-12,08;12,56)	1,96 (-17,21;23,38)	2,02 (-3,04;6,70)
biais relatif	0,02	-0,01	-0,02	0,01
écart type <sup>1</sup>	13,59	3,29	6,15	1,52
écart type <sup>2</sup>	13,07	3,26	6,05	1,49
écart type <sup>3</sup>	14,70	3,25	6,72	1,47
beta1 (min,max)	4,00 (2,27;5,64)	4,00 (3,49;4,53)	4,00 (3,14;4,75)	4,00 (3,78;4,24)
écart type <sup>1</sup>	0,52	0,14	0,24	0,07
écart type <sup>2</sup>	0,50	0,14	0,23	0,07
écart type <sup>3</sup>	0,56	0,14	0,26	0,06
beta2 (min,max)	-3,00 (-4,35;-1,41)	-3,00 (-3,65;-2,30)	-3,00 (-3,59;-2,34)	-3,00 (-3,27;-2,76)
écart type <sup>1</sup>	0,44	0,18	0,20	0,08
écart type <sup>2</sup>	0,42	0,17	0,20	0,08
écart type <sup>3</sup>	0,47	0,18	0,22	0,08
beta3 (min,max)	-5,00 (-5,90;-4,23)	-5,00 (-5,28;-4,72)	-5,00 (-5,38;-4,62)	-5,00 (-5,14;-4,87)
écart type <sup>1</sup>	0,25	0,08	0,11	0,04
écart type <sup>2</sup>	0,24	0,08	0,11	0,04
écart type <sup>3</sup>	0,26	0,08	0,12	0,04
beta4 (min,max)	9,00 (7,66;10,26)	9,00 (8,39;9,53)	9,00 (8,42;9,63)	9,00 (8,68;9,25)
écart type <sup>1</sup>	0,39	0,16	0,18	0,08
écart type <sup>2</sup>	0,38	0,16	0,17	0,07
écart type <sup>3</sup>	0,43	0,17	0,19	0,08

<sup>1</sup> écart type obtenu lors des simulations

<sup>2</sup> écart type obtenu avec l'estimateur de Liang et Zeger (1986)

<sup>3</sup> écart type obtenu avec l'estimateur de Prentice (1988)

TABLEAU 8.2

Corrélation des observations: AR(1)

Corrélation de départ: Indépendante

Niveau de corrélation: 0,45

Type de données: Normale

	# de sujets:=20		# de sujets:=100	
	# de répétitions par sujet:=5	# de répétitions par sujet:=10	# de répétitions par sujet:=5	# de répétitions par sujet:=10
beta0 (min,max)	2,01 (-27,52;31,76)	1,98 (-10,28;12,97)	2,01 (-16,27;23,04)	2,01 (-3,02;7,56)
biais relatif	0,01	-0,01	0,01	0,01
écart type <sup>1</sup>	9,86	3,41	4,56	1,56
écart type <sup>2</sup>	9,58	3,37	4,55	1,55
écart type <sup>3</sup>	14,69	3,24	4,47	1,47
beta1 (min,max)	4,00 (2,81;5,24)	4,00 (3,48;4,51)	4,00 (3,12;4,71)	4,00 (3,75;4,22)
écart type <sup>1</sup>	0,38	0,15	0,17	0,07
écart type <sup>2</sup>	0,37	0,15	0,17	0,07
écart type <sup>3</sup>	0,56	0,14	0,16	0,06
beta2 (min,max)	-3,00 (-3,97;-2,02)	-3,00 (-3,58;-2,37)	-3,00 (-3,63;-2,37)	-3,00 (-3,25;-2,78)
écart type <sup>1</sup>	0,33	0,16	0,17	0,07
écart type <sup>2</sup>	0,32	0,15	0,17	0,07
écart type <sup>3</sup>	0,47	0,18	0,18	0,08
beta3 (min,max)	-5,00 (-5,85;-4,37)	-5,00 (-5,31;-4,68)	-5,00 (-5,35;-4,65)	-5,00 (-5,15;-4,85)
écart type <sup>1</sup>	0,21	0,09	0,07	0,04
écart type <sup>2</sup>	0,20	0,09	0,07	0,04
écart type <sup>3</sup>	0,26	0,08	0,08	0,04
beta4 (min,max)	9,00 (7,87;9,83)	9,00 (8,44;9,48)	9,00 (8,47;9,55)	9,00 (8,73;9,20)
écart type <sup>1</sup>	0,28	0,14	0,16	0,06
écart type <sup>2</sup>	0,27	0,14	0,16	0,06
écart type <sup>3</sup>	0,43	0,17	0,18	0,08

<sup>1</sup> écart type obtenu lors des simulations

<sup>2</sup> écart type obtenu avec l'estimateur de Liang et Zeger (1986)

<sup>3</sup> écart type obtenu avec l'estimateur de Prentice (1988)

TABLEAU 8.3

Corrélation des observations: AR(1)

Corrélation de départ: Indépendante

Niveau de corrélation: 0,15

Type de données: Binaire

	# de sujets:=20		# de sujets:=100	
	# de répétitions par sujet:=5	# de répétitions par sujet:=10	# de répétitions par sujet:=5	# de répétitions par sujet:=10
beta0 (min,max)	0,06 (-9,54;8,99)	0,01 (-1,11;1,05)	-0,00 (-1,38;1,42)	0,00 (-0,35;0,38)
biais relatif	0,06	0,01	0,00	0,00
écart type <sup>1</sup>	3,09	0,31	0,41	0,1
écart type <sup>2</sup>	2,07	0,46	0,84	0,20
écart type <sup>3</sup>	2,35	0,46	0,96	0,20
beta1 (min,max)	0,12 (-4,81;5,18)	0,14 (-0,35;0,67)	0,14 (-0,48;0,85)	0,14 (-0,04;0,28)
écart type <sup>1</sup>	1,57	0,19	0,21	0,05
écart type <sup>2</sup>	1,06	0,22	0,43	0,10
écart type <sup>3</sup>	1,18	0,21	0,49	0,10
beta2 (min,max)	-0,09 (-3,30;3,76)	-0,10 (-0,40;0,24)	-0,10 (-0,49;0,36)	-0,10 (-0,18;-0,04)
écart type <sup>1</sup>	1,2	0,11	0,13	0,02
écart type <sup>2</sup>	0,79	0,12	0,32	0,05
écart type <sup>3</sup>	0,85	0,12	0,35	0,05
beta3 (min,max)	-0,91 (-2,76;0,54)	-0,87 (-1,63;-0,35)	-0,85 (-1,22;-0,45)	-0,85 (-1,14;-0,65)
écart type <sup>1</sup>	0,53	0,23	0,12	0,07
écart type <sup>2</sup>	0,56	0,29	0,24	0,13
écart type <sup>3</sup>	0,62	0,33	0,27	0,15
beta4 (min,max)	1,05 (-10,44;11,33)	1,11 (-1,43;3,58)	1,09 (0,05;2,47)	1,10 (0,43;1,87)
écart type <sup>1</sup>	3,5	0,8	0,35	0,23
écart type <sup>2</sup>	2,27	0,93	0,93	0,42
écart type <sup>3</sup>	2,29	0,91	0,93	0,40

<sup>1</sup> écart type obtenu lors des simulations

<sup>2</sup> écart type obtenu avec l'estimateur de Liang et Zeger (1986)

<sup>3</sup> écart type obtenu avec l'estimateur de Prentice (1988)

TABLEAU 8.4

Corrélation des observations: AR(1)

Corrélation de départ: Indépendante

Niveau de corrélation: 0,30

Type de données: Binaire

	# de sujets:=20		# de sujets:=100	
	# de répétitions par sujet:=5	# de répétitions par sujet:=10	# de répétitions par sujet:=5	# de répétitions par sujet:=10
beta0 (min,max)	-0,03 (-8,11;6,07)	-0,01 (-1,34;1,17)	-0,01 (-1,77;1,54)	0,00 (-0,40;0,40)
biais relatif	-0,03	-0,01	-0,01	0,00
écart type <sup>1</sup>	2,34	0,43	0,48	0,13
écart type <sup>2</sup>	1,80	0,47	0,70	0,21
écart type <sup>3</sup>	2,36	0,46	0,96	0,20
beta1 (min,max)	0,16 (-3,43;4,70)	0,15 (-0,57;1,00)	0,14 (-0,60;1,05)	0,14 (-0,04;0,32)
écart type <sup>1</sup>	1,21	0,25	0,24	0,06
écart type <sup>2</sup>	,95	0,23	0,37	0,10
écart type <sup>3</sup>	1,18	0,21	0,48	0,09
beta2 (min,max)	-0,08 (-2,85;3,41)	-0,11 (-0,49;0,33)	-0,10 (-0,59;0,52)	-0,10 (-0,21;0,00)
écart type <sup>1</sup>	1,09	0,13	0,18	0,03
écart type <sup>2</sup>	0,74	0,13	0,29	0,06
écart type <sup>3</sup>	0,86	0,12	0,35	0,05
beta3 (min,max)	-0,92 (-3,26;0,37)	-0,87 (-1,59;-0,11)	-0,85 (-1,42;-0,23)	-0,85 (-1,19;-0,62)
écart type <sup>1</sup>	0,49	0,23	0,18	0,09
écart type <sup>2</sup>	0,50	0,26	0,21	0,12
écart type <sup>3</sup>	0,63	0,33	0,27	0,15
beta4 (min,max)	1,08 (-10,18;9,74)	1,14 (-1,68;4,16)	1,09 (-0,44;2,74)	1,10 (0,18;2,04)
écart type <sup>1</sup>	3,52	1,04	0,54	0,28
écart type <sup>2</sup>	2,31	0,97	0,93	0,43
écart type <sup>3</sup>	2,30	0,92	0,93	0,40

<sup>1</sup> écart type obtenu lors des simulations

<sup>2</sup> écart type obtenu avec l'estimateur de Liang et Zeger (1986)

<sup>3</sup> écart type obtenu avec l'estimateur de Prentice (1988)

## BIBLIOGRAPHIE

---

DUNLOP, D.D., *Regression for longitudinal data: A bridge from least squares regression.*, The American Statistician, 48(4), 299-303, 1994.

EMRICH, L.J. & PIEDMONTE, M.R., *A method for generating high-dimensional multivariate binary variates.*, The American Statistician, 45(4), 302-305, 1991

EMRICH, L.J. & PIEDMONTE, M.R., *On some small sample properties of generalized estimating equation estimates for multivariate dichomomous outcomes.*, J. Statist. Comput. Simul., 41, 19-29, 1992.

GUNSOLLEY, J.C., GETCHELL, C. & CHINCHILLI, V.M., *Small sample characteristics of generalized estimating equations.*, Commun. Statist.-Simula., 24(4), 869-878, 1995.

KARIM, M.R. & ZEGER, S.L., *GEE: A SAS Macro for Longitudinal Data Analysis.*, Technical report 674, The John Hopkins University, Dept. of Biostatistics, 1988.

LIANG, K.Y. & ZEGER, S.L., *Longitudinal data analysis using generalized linear models.*, Biometrika, 73, 13-22, 1986.

LIPSITZ, S.R., FITZMAURICE, M., ORAV, E.J. & LIARD, N.M., *Performance of generalized estimating equations in practical situations.*, Biometrics, 50, 270-278, 1994.

PAIK, M.C., *Repeated measurement analysis for nonnormal data in small samples.*, Commun. Statist.-Simula., 17(4), 1155-1171, 1988.

PRENTICE, R.L., *Correlated binary regression with covariates specific to each binary observation.*, Biometrics, 44, 1033-1048, 1988.

SAS INSTITUTE INC. (1990), *SAS Procedures Guides.* version 6 third edition. SAS Institut, Inc. Cary, NC, USA.

SHARPLES, K. & BRESLOW, N., *Regression analysis of correlated binary data: Some small sample results for the estimating equation approach.*, J. Statist. Comput. Simul. , 42, 1-20, 1992.

ZEGER, S.L. & LIANG, K.Y., *Longitudinal data analysis for discrete and continuous outcomes.*, Biometrics, 42, 121-130, 1986.