GLMM

Plan

1. Introduction

- Concept introduit par Nelder & Wedderburn (1972)
- Etendre le modèle linéaire à des données non gaussiennes
- Développement par
  - Mc Cullagh & Nelder (cf ouvrage, 1987)
  - Liang & Zeger (1986): GEE
- Applications nombreuses: variables discrètes et continues

2. Fonction de lien

$a) y_i \sim_{id} \left(\mu_i, \sigma_i^2\right)$ famille **exponentielle** simple

Binomiale, Poisson; Normale, Gamma

$b)$ Modèle linéaire: $E\left(y_i\right) = \mu_i = \mathbf{x}_i^{'}\boldsymbol{\beta}$

GLM : $\boxed{\eta_i = g\left(\mu_i\right) = \mathbf{x}_i^{'}\boldsymbol{\beta}}$; $g =$ **fonction de lien**

fonction inverse: $\mu_i = h\left(\eta_i\right)$

$Ex : \eta = \mu$: identité; $\eta = \ln \mu$: logarithme (Poisson)

$\eta = \ln \frac{\mu}{1-\mu}$: logit ; $\eta = \Phi^{-1}\left(\mu\right)$: Probit (Binomiale)

2. Famille exponentielle

Y est distribuée dans la famille exponentielle simple

si sa densité (ou fonction de probabilité) peut s'écrire

sous la forme:

$$f_Y(y;\theta,\phi) = \exp\left[\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\right]$$

$a(\phi), b(\theta)$ et $c(y,\phi)$ = fonctions spécifiques de chaque distribution

θ : paramètre naturel; ϕ : paramètre d'échelle ou de dispersion

ϕ connu $\Rightarrow$ distribution dépend d'un seul paramètre

(binomiale, Poisson; ϕ=1)

ϕ inconnu $\Rightarrow$ f.exponentielle à 2 paramètres (normale, gamma) ; $a(\phi) = \phi / w$

2. Famille exponentielle/Binomiale

$$\ln f_Y\left(y;\theta,\phi\right)=\frac{y\theta-b\left(\theta\right)}{a\left(\phi\right)}+c\left(y,\phi\right);va\ fréquence: Y=Z/n$$

$$\Pr(Y=y)=C_n^{ny}\pi^{ny}\left(1-\pi\right)^{n(1-y)}$$

$$\ln\Pr(Y=y)=ny\ln\pi+n(1-y)\ln\left(1-\pi\right)+\ln C_n^{ny}$$

$$\ln\Pr(Y=y)=ny\left[\ln\pi-\ln\left(1-\pi\right)\right]+n\ln\left(1-\pi\right)+\ln C_n^{ny}$$

$$\ln\Pr(Y=y)=\underbrace{n}_{\frac{1}{a(\phi)}}\left[y\underbrace{\ln\frac{\pi}{1-\pi}}_{\theta}+\underbrace{\ln\left(1-\pi\right)}_{-b(\theta)}\right]+\underbrace{\ln C_n^{ny}}_{c(y,\phi)}$$

$$\theta=\ln\frac{\pi}{1-\pi}\Rightarrow e^{\theta}=\frac{\pi}{1-\pi}\Rightarrow\pi=\frac{e^{\theta}}{1+e^{\theta}}=\frac{1}{1+e^{-\theta}};\phi=1,w=n$$

2. Famille exponentielle/Moments

$$\boxed{E(y) = \mu = b'(\theta)}$$

$$Var(y) = \underbrace{a(\phi)}_{\frac{\phi}{w}} b''(\theta) = \phi \frac{b''(\theta)}{w} = \phi v(\mu)$$

$$\boxed{Var(y) = \phi v(\mu)}; \; v(\mu) = \frac{1}{w}\frac{\partial \mu}{\partial \theta} = f^{on} \text{ de variance}$$

2. Fonction de lien

c)Lien **canonique**

θ: paramètre naturel; ϕ: paramètre d'échelle

L. canonique = fonction $\theta_i = \theta(\mu_i)$

avantage: stat exhaustive pour θ_i

2. Famille exponentielle/caractéristiques

	Normal	*Poisson*	*Binomial*	*Gamma*	*Inverse Gaussian*
Notation	$N(\mu,\sigma^2)$	$P(\mu)$	$B(m,\pi)/m$	$G(\mu,\nu)$	$IG(\mu,\sigma^2)$
Range of y	$(-\infty,\infty)$	$0(1)\infty$	$\frac{0(1)m}{m}$	$(0,\infty)$	$(0,\infty)$
Dispersion parameter: ϕ	$\phi=\sigma^2$	1	$1/m$	$\phi=\nu^{-1}$	$\phi=\sigma^2$
Cumulant function: $b(\theta)$	$\theta^2/2$	$\exp(\theta)$	$\log(1+e^\theta)$	$-\log(-\theta)$	$-(-2\theta)^{1/2}$
$c(y;\phi)$	$-\frac{1}{2}\left(\frac{y^2}{\phi}+\log(2\pi\phi)\right)$	$-\log y!$	$\log\binom{m}{my}$	$\nu\log(\nu y)-\log y$ $-\log\Gamma(\nu)$	$-\frac{1}{2}\left\{\log(2\pi\phi y^3)+\frac{1}{\phi y}\right\}$
$\mu(\theta)=E(Y;\theta)$	θ	$\exp(\theta)$	$e^\theta/(1+e^\theta)$	$-1/\theta$	$(-2\theta)^{-1/2}$
Canonical link: $\theta(\mu)$	identity	log	logit	reciprocal	$1/\mu^2$
Variance function: $V(\mu)$	1	μ	$\mu(1-\mu)$	μ^2	μ^3

†The mean-value parameter is denoted by μ, or by π for the binomial distribution.
The parameterization of the gamma distribution is such that its variance is μ^2/ν.
The canonical parameter, denoted by θ, is defined by (2.4). The relationship between μ and θ is given in lines 6 and 7 of the Table.

2. Estimation ML

$$\mathbf{y}_{N\text{x}1} = \left\{ y_i \right\}, y_i \sim_{id} \text{ famille exponentielle simple}$$

$$\mathrm{E}\left(\mathbf{y}\right) = \boldsymbol{\mu};\ \mathrm{Var}(\mathbf{y}) = \boldsymbol{\Sigma} = \phi\mathbf{V};\ \mathbf{V} = \left\{ v\left(\mu_i\right) \right\}$$

$$\text{Score} : \mathrm{U}\left(\boldsymbol{\beta}\right) = \frac{\partial \mathrm{L}\left(\boldsymbol{\beta};\mathbf{y}\right)}{\partial \boldsymbol{\beta}} = \sum_i \mathrm{U}_i\left(\boldsymbol{\beta}\right)$$

2. Estimation ML

$$\boxed{\mathbf{U} = \mathbf{D}'\boldsymbol{\Sigma}^{-1}\left(\mathbf{y} - \boldsymbol{\mu}\right)} \quad \mathbf{U} = 0 \Rightarrow \hat{\boldsymbol{\beta}}_{ML}$$

$$\mathbf{D}' = \mathbf{X}'\mathbf{K};\ \mathbf{K}_{NxN} = Diag\left\{k_i = \frac{\partial \mu_i}{\partial \eta_i}\right\}$$

$$\text{Rappel : M.linéaire } \mathbf{X}'\boldsymbol{\Sigma}^{-1}\left(\mathbf{y} - \underbrace{\boldsymbol{\mu}}_{\mathbf{X}\boldsymbol{\beta}}\right) = \mathbf{0}$$

2. Estimation ML/Moindres carrés pondérés itérés

$\mathbf{D}'\mathbf{\Sigma}^{-1}\left(\mathbf{y}-\boldsymbol{\mu}\right)=0$ Résolution Fisher

En pratique, résoudre **itérativement**

le syst de moindres carrés pondérés

$$\boxed{\mathbf{X}'\mathbf{W}^{[t]}\mathbf{X}\boldsymbol{\beta}^{[t+1]}=\mathbf{X}'\mathbf{W}^{[t]}\mathbf{z}^{[t]}}$$

$$\mathbf{W}=\mathbf{K}\mathbf{\Sigma}^{-1}\mathbf{K}=\mathbf{K}\mathbf{V}^{-1}\mathbf{K}/\phi$$

$$\mathbf{K}_{NxN}=Diag\left\{k_i=\partial\mu_i/\partial\eta_i\right\}$$

$$\mathbf{z}^{[t]}=\mathbf{X}\boldsymbol{\beta}^{[t]}+\left(\mathbf{K}^{[t]}\right)^{-1}\left(\mathbf{y}-\boldsymbol{\mu}^{[t]}\right)$$

Existence ML: Séparabilité en analyse binaire

2. Surdispersion

Binomiale et Poisson : ϕ=1

Déviance (D) ou X_g^2 Pearson trop grands (**surdispersion**)

Rappel: D*=D/ϕ ou $X^2 = X_g^2/\phi \rightarrow Khi^2$

D et $X_g^2 \gg$ nbre de degrés de liberté

Causes:

1) Echantillonnage des probabilités de réponse

Ex: modèle beta-binomial; Poisson-gamma (=binomiale négative)

2)Corrélation entre observations d'un meme sous groupe

$$Var(y_{i+}) = n_i \pi_i (1-\pi_i)\left[1+(n_i-1)\rho\right]$$

2. Surdispersion: quasi-score

$$\mathbf{D}'\boldsymbol{\Sigma}^{-1}\left(\mathbf{y}-\boldsymbol{\mu}\right)=\mathbf{D}'\mathbf{V}^{-1}\left(\mathbf{y}-\boldsymbol{\mu}\right)/\phi$$

$$\mathbf{W}=\mathbf{K}\boldsymbol{\Sigma}^{-1}\mathbf{K}=\mathbf{K}\mathbf{V}^{-1}\mathbf{K}/\phi=\underline{\mathbf{W}}/\phi$$

$$\boxed{\mathbf{X}'\underline{\mathbf{W}}^{[t]}\mathbf{X}\boldsymbol{\beta}^{[t+1]}=\mathbf{X}'\underline{\mathbf{W}}^{[t]}\mathbf{z}^{[t]}}$$

$\hat{\boldsymbol{\beta}}$ ne dépend pas de ϕ contrairement à $\mathrm{Var}_{\mathrm{as}}(\hat{\boldsymbol{\beta}})=\phi\left(\mathbf{X}'\hat{\underline{\mathbf{W}}}\mathbf{X}\right)^{-1}$

si ϕ inconnu Quasi-scores : $\boxed{\mathbf{D}'\mathbf{V}^{-1}\left(\mathbf{y}-\boldsymbol{\mu}\right)=\mathbf{0}}$

ϕ (surdispersion) estimé par

$$\hat{\phi}=X_g^2/\left[N-r(\mathbf{X})\right];X_g^2=\sum_{i=1}^{N}\frac{\left(y_i-\hat{\mu}_i\right)^2}{v(\hat{\mu}_i)}\text{ (Stat de Pearson gen.)}$$

$$\hat{\phi}=\text{Deviance}/\left[N-r(\mathbf{X})\right];\text{Deviance}/\phi=-2L(y;\hat{\beta})+2L(y;y)$$

3. Données corrélées/modélisation

- Modèle marginal: « Population average »
- Modèle mixte: « Subject specific »

3. Données corrélées/Modèles

1) Modèle marginal   2) Modèle mixte

En modèle linéaire: 1) equivalent à 2)

$$\mathrm{E}\left(y_i\right)=\mathrm{E}_{\mathbf{u}}\left[\mathrm{E}\left(y_i \mid \mathbf{u}\right)\right]=\mathrm{E}_{\mathbf{u}}\left[\mathbf{x}_i^{'}\boldsymbol{\beta}+z_i^{'}\mathbf{u}\right]=\mathbf{x}_i^{'}\boldsymbol{\beta}$$

En modèle non linéaire: faux

$$\mathrm{E}\left(y_i\right)=\mathrm{E}_{\mathbf{u}}\left[\mathrm{E}\left(y_i \mid \mathbf{u}\right)\right]$$

$$=\mathrm{E}_{\mathbf{u}}\left[h(\mathbf{x}_i^{'}\boldsymbol{\beta}+z_i^{'}\mathbf{u})\right]\neq h\left[\underbrace{\mathrm{E}_{\mathbf{u}}(\mathbf{x}_i^{'}\boldsymbol{\beta}+z_i^{'}\mathbf{u})}_{\left(\mathbf{x}_i^{'}\boldsymbol{\beta}\right)}\right]$$

3. Données corrélées/Modèle marginal

$$\mathrm{E}\left(y_i\right)=\mu_i=h\left(\eta_\mathrm{i}\right);\ \eta_\mathrm{i}=\mathbf{x}_i^{'}\boldsymbol{\beta}$$

$$\mathrm{Var}(y_i)=\sigma_i^2;\ \mathrm{Cov}\left(y_i,y_{i'}\right)=\sigma_{ii'}$$

σ_i^2 et $\sigma_{ii'}$ fonctions de $\boldsymbol{\beta}$ et autres par.

3. Données corrélées/Modèle mixte

Modèle hiérarchique

$$1)\,\mathrm{E}\left(y_i \mid \mathbf{u}\right) = \mu_{u,i} = h\left(\eta_{u,i}\right) \quad \eta_{u,i} = \mathbf{x}_i^{'}\boldsymbol{\beta} + \mathbf{z}_i^{'}\mathbf{u}$$

$$\mathrm{Var}\left(y_i \mid \mathbf{u}\right) = \phi v\left(\mu_{u,i}\right)$$

$$2)\,\mathbf{u} \sim \mathcal{N}\left(\mathbf{0}, \mathbf{G}\right)$$

3. Données corrélées/Exemple binomial

$$y_{ij} \mid \mathbf{u} \sim B\left(1, \pi_{u,i}\right); y_{i+} \mid \mathbf{u} \sim B\left(n_i, \pi_{u,i}\right)$$

$$\pi_{u,i} = \Phi\left(\eta_{u,i}\right) \Leftrightarrow \eta_{u,i} = \Phi^{-1}\left(\pi_{u,i}\right);\ \eta_{u,i} = \mathbf{x}_i^{'}\boldsymbol{\beta} + \mathbf{z}_i^{'}\mathbf{u}$$

$$\pi_i = \mathrm{E}\left(y_{ij}\right) = \mathrm{E}_{\mathbf{u}}\left[\mathrm{E}\left(y_{ij} \mid \mathbf{u}\right)\right] = \mathrm{E}_{\mathbf{u}}\left[\Phi\left(\eta_{u,i}\right)\right]$$

$$\mathrm{E}_{\mathbf{u}}\left[\Phi\left(\mathbf{x}_i^{'}\boldsymbol{\beta} + \mathbf{z}_i^{'}\mathbf{u}\right)\right] = \Phi\left(\frac{\mathbf{x}_i^{'}\boldsymbol{\beta}}{\sqrt{1 + \mathbf{z}_i^{'}\mathbf{G}\mathbf{z}_i}}\right)$$

$$\text{Si } \mathbf{z}_i^{'}\mathbf{G}\mathbf{z}_i = Cste = \sigma_u^2 \Rightarrow \pi_i = \Phi\left(\frac{\mathbf{x}_i^{'}\boldsymbol{\beta}}{\sqrt{1+\sigma_u^2}}\right)$$

$$\text{Modèle marginal :}\ \pi_i = \Phi\left(\mathbf{x}_i^{'}\boldsymbol{\beta}^M\right)$$

$\boldsymbol{\beta}^M$ M.marginal (Pop average) $\neq \boldsymbol{\beta}^{RE}$ M.mixte (Sub. specific)

$$\text{Probit :}\ \boldsymbol{\beta}^{RE} / \boldsymbol{\beta}^M = \sqrt{1+\sigma_u^2};$$

$$\text{Logit :}\ \boldsymbol{\beta}^{RE} / \boldsymbol{\beta}^M \approx \sqrt{1+\left(\sigma_u^2 / c^2\right)}; c = (15/16)\pi/\sqrt{3} = 1.7$$

4. Approche GEE

GEE=Modèle **marginal** : Liang&Zeger, 1986, Biometrika, 73,13-22

Données longitudinales: $\mathbf{y}_i = \left(y_{i1}, y_{i2}, ..., y_{it}, ..., \right)'$

$$1)\, \mathrm{E}\left(y_{it} \right) = \mu_{it} = h\left(\mathbf{x}_{it}^{'} \boldsymbol{\beta} \right);\ 2) \mathrm{Var}\left(\mathbf{y}_i \right) = \boldsymbol{\Sigma}_i = \phi \mathbf{V}_i \left(\alpha, \boldsymbol{\beta} \right)$$

$$.\mathrm{Var}\left(y_{it} \right) = \phi v\left(\mu_{it} \right)$$

$$.\mathrm{Corr}(y_{is}, y_{it}) = r_{i,st}\left(\alpha \right) \text{ "working correlation"}$$

$$\mathbf{V}_i\left(\alpha, \boldsymbol{\beta} \right) = \left[\mathbf{A}_i\left(\boldsymbol{\beta} \right) \right]^{1/2} \mathbf{R}_i\left(\alpha \right) \left[\mathbf{A}_i\left(\boldsymbol{\beta} \right) \right]^{1/2}$$

$$\left[\mathbf{A}_i\left(\boldsymbol{\beta} \right) \right] = \mathrm{Diag}\left\{ v\left(\mu_{it} \right) \right\}\ Ex:\ \pi_{it}\left(1 - \pi_{it} \right)$$

$$r_{i,st}\left(\alpha \right) = 0;\ r_{i,st}\left(\alpha \right) = \alpha;\ r_{i,st}\left(\alpha \right) = \alpha^{|s-t|}$$

4. GEE/Estimation

Quasi-Score : $\mathbf{D}'\mathbf{V}^{-1}\left(\mathbf{y}-\boldsymbol{\mu}\right)=0$

GEE=Extension quasi-score ie $\mathbf{V}=\mathbf{V}\left(\boldsymbol{\beta},\hat{\alpha}\right)$

$\hat{\alpha}=\hat{\alpha}\left(\mathbf{y},\boldsymbol{\beta},\phi\right)$est. convergent de α qd $\boldsymbol{\beta},\phi$ connus

En pratique, on remplace ϕ par est converg. $\hat{\phi}\left(\mathbf{y},\boldsymbol{\beta}\right)$

Résultat : $\sqrt{N}\left(\hat{\boldsymbol{\beta}}_G-\boldsymbol{\beta}\right)\rightarrow N\left(0,\mathbf{V}_G\right)$

$\mathbf{V}_G=Lim N\mathbf{H}_0^{-1}\mathbf{H}_1\mathbf{H}_0^{-1}$(est sandwich)

$\mathbf{H}_0=\sum_i \mathbf{D}_i'\mathbf{V}_i^{-1}\mathbf{D}_i;\mathbf{H}_1=\sum_i \mathbf{D}_i'\mathbf{V}_i^{-1}\boldsymbol{\Sigma}_i\mathbf{V}_i^{-1}\mathbf{D}_i$

$\boldsymbol{\Sigma}_i$ remplacé par un est. convergent: $\hat{\boldsymbol{\Sigma}}_i=\left(\mathbf{y}_i-\hat{\boldsymbol{\mu}}_i\right)\left(\mathbf{y}_i-\hat{\boldsymbol{\mu}}_i\right)'$

Test robuste (mauvaise spécification de $\boldsymbol{\Sigma}_i$)

5. Modèle mixte/Inférence

– Approche par la vraisemblance
  – intégration numérique (Anderson & Aitkin, 1985) et/ou EM (Zeger & Karim, 1991; Booth & Albert, 1999; Mc Culloch,1997; Levine & Casella, 2001)
– Approche bayésienne
  – MAP (Gianola & Foulley, 1983; Harville & Mee,1984; Laird & Stiratelli, 1984)
  – Procédés dérivés (PQL, Breslow & Clayton, 1993; Schall, 1991)
  – Complète: MCMC (Albert & Chib, 1993)
– Approche par vraisemblance hiérarchique (h-likelihood)
  – Lee & Nelder (1996)

6. Offre logicielle

- **Offre abondante**
  - **SAS**
  - S-Plus
  - GLIM
  - GENSTAT
  - ASRELM
  - SPSS
  - **WINBUGS**
- **Profusion de méthodes**

6. Offre logicielle/SAS

Procédures SAS® pour les GLM et GLMM

Procédure	Données	Lien	Modèle	Méthode
Logistic	Binaire Ordinales	Logit Probit Cloglog	Effets Fixes, GLM	ML
Catmod	Binaires, Nominales	Logit Loglinéaire	Effets fixes	ML (logit, loglinéaire) WLS*(GLM)
Genmod	Binaires Ordinales Continues Comptage	Logit, Probit Log, Cloglog Puissance	GLM, Modèle Marginal Modèle conditionnel ?	ML (fixe ind) GEE § (marginal) Exacte (m.conditionnel)
Nlmixed	Quelconques	Quelconque	EffetsMixtes	ML (Quadrature)
Glimmix£				approchée

*Grizzle, Starmer & Koch, 1969

§Liang & Zeger, 1986 ; Zeger & Liang, 1986; £Macro

6. Exemple/données « orobanche »

Data set: Germination of seeds of two species using 2 extracts in different batches (Crowder, 1978; Collett, 2002)

O Aegyptiaca 75				O Aegyptiaca 73			
Bean		Cucumber		Bean		Cucumber	
Y	N	Y	N	Y	N	Y	N
10	39	5	6	8	16	3	12
23	62	53	74	10	30	22	41
23	81	55	72	8	28	15	30
26	51	32	51	23	45	32	51
17	39	46	79	0	4	3	7
		10	13				

6. Exemple/orobanche/modèle

$$\text{Additive/logit}\left(\pi_{ijk}\right) = \text{constant} + species_i + extract_j + \gamma z_{ijk}$$

π_{ijk} : true response probability in batch k of species i and extract j

$$\text{Interactive/logit}\left(\pi_{ijk}\right) = \text{constant} + species_i + extract_j + \left(se\right)_{ij} + \gamma z_{ijk}$$

$z_{ijk} \sim \mathcal{N}\left(0,1\right)$: k^{th} random batch effect (standardized) within $\text{i}^{\text{th species}}$ and j^{th} extract

6. Exemple/ orobanche/résultats

	GLM	GLM Overdisp.	PQL	MQL	ML Nlmixed, Egret	Post Mean Winbugs
Var.	est±SE	est±SE	est±SE	est±SE	est±SE	est±SE
			Additive models			
Constant	-.430±.114	-.430±.169	-.375±.182	-.369±.180	-.389±.165	-.389±.189
Species	-.270±.155	-.270±.230	-.363±.228	-.357±.227	-.347±.215	-.360±.237
Extract	1.065±.144	1.065±.214	1.012±.224	.998±.222	1.029±.205	1.033±.232
Phi*.5	1	1.4848			1	
gamma			.352±.118	.349±117	.295±.112	.338±.147
IC					aic=119.7	dic=114.1
			Interaction models			
Constant	-.558±.126	-.558±.176	-.542±.190	-.536±.190	-.548±.167	-.556±.191
Species	.146±.223	.146±.312	.077±.308	.074±.308	097±.278	.088±.309
Extract	1.318±.178	1.318±.248	1.339±.270	1.326±.269	1.337±.237	1.358±.271
SxE	-.778±.306	-.778±.429	-.825±.430	-.816±.429	-.810±.385	-.829±.428
Test SE	6.45 (0.011)	3.29 (0.070)	3.68 (0.055)	3.62 (0.057)	4.43 (0.048)	3.75 (0.053)
Phi*.5	1	1.3991			1	
gamma			.313±.121	.313±.120	.236±.110	.280±.142
IC					aic=117.5	dic=113.3

*Informations about PQL, MQL from Breslow & Clayton (1993) and about ML Egret from Collett (1991)

Constant=Species(75) with Extract(Bean)
Additive Model: Species=73-75 ; Extract=Cucumber-Bean
Interactive Model: Species=73-75 with Bean; Extract=Cucumber-Bean with 75

6. Exemple/ orobanche/résultats

Test de l'interaction et de l'effet aléatoire « coupelle »

	GLM	Mixte	Δ(-2L)	P-value[$]
Additif	116.3	111.7	4.6	0.016
Interaction	109.9	107.5	2.4	0.061
Δ(-2L)	6.4	4.2		
P-value*	0.011	0.040		

-2L=Moins deux fois la logvraisemblance max

Δ(-2L) : variation de -2L ; GLM-Mixte et Additif-Interaction

*Distribution asymptotique : Khi2 à 1 degré de liberté

[$]Distribution asymptotique : Mélange 1/2Dirac(0)+1/2 Khi2 à 1 degré de liberté

6. Exemple «orobanche»/fichier

Plate	Spec	Extr	Germ	Count
1	75	EB	10	39
2	75	EB	23	62
3	75	EB	23	81
4	75	EB	26	51
5	75	EB	17	39
6	75	C	5	6
7	75	C	53	74
8	75	C	55	72
9	75	C	32	51
10	75	C	46	79
11	75	C	10	13
12	73	EB	8	16
13	73	EB	10	30
14	73	EB	8	28
15	73	EB	23	45
16	73	EB	0	4
17	73	C	3	12
18	73	C	22	41
19	73	C	15	30
20	73	C	32	51
21	73	C	3	7

6. Exemple «orobanche»/genmod

```
proc genmod data=oro;
   title germination SPEC EXTR SPEC*EXTR Fixe phi=1 ML;
   class spec extr;
   model  germ/count=spec extr spec*extr/dist=bin type3 wald ;
 run;
    proc genmod data=oro;
   title germination SPEC EXTR SPEC*EXTR Fixe over phie GLM;
   class spec extr;
   model  germ/count=spec extr spec*extr/dist=bin type3 wald scale=deviance;
 run;
```

Analysis Of Parameter Estimates

Parameter			DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr
Intercept			1	-0.5582	0.1763	-0.9037	-0.2126	10.02	
Spec	73		1	0.1459	0.3122	-0.4660	0.7579	0.22	
Spec	75		0	0.0000	0.0000	0.0000	0.0000	.	
Extr	C		1	1.3182	0.2483	0.8315	1.8048	28.18	
Extr	EB		0	0.0000	0.0000	0.0000	0.0000	.	
Spec*Extr	73	C	1	-0.7781	0.4287	-1.6184	0.0622	3.29	
Spec*Extr	73	EB	0	0.0000	0.0000	0.0000	0.0000	.	
Spec*Extr	75	C	0	0.0000	0.0000	0.0000	0.0000	.	
Spec*Extr	75	EB	0	0.0000	0.0000	0.0000	0.0000	.	
Scale			0	1.3991	0.0000	1.3991	1.3991		

6. Exemple/données « epilepsie »/données

Table 8.6 *Occurrence of seisures in four successive two-week periods.*

Individual	Y_1	Y_2	Y_3	Y_4	*Treat*	*Age*	*Count*
1	1	0	0	0	0	31	11
2	0	1	0	0	0	30	11
3	0	0	0	1	0	25	6
4	0	0	0	0	0	36	8
5	1	1	1	1	0	22	66
6	1	0	1	1	0	29	27
7	1	0	0	0	0	31	12
8	1	1	1	1	0	42	52
9	1	1	1	1	0	37	23
10	1	1	1	0	0	28	10
11	1	1	1	1	0	36	52
12	1	1	1	0	0	24	33
13	0	0	1	0	0	23	18
14	1	1	1	1	0	36	42
15	1	1	1	1	0	26	87
16	1	0	0	1	0	26	50
17	0	0	0	0	0	28	18
18	1	1	1	1	0	31	111
19	0	1	0	1	0	32	18
20	0	0	1	1	0	21	20
21	0	0	0	0	0	29	12
22	0	0	0	0	0	21	9
23	0	0	0	1	0	32	17
24	1	1	0	1	0	25	28
25	1	1	1	1	0	30	55
26	0	0	0	0	0	40	9
27	0	0	0	0	0	19	10
28	1	1	1	1	0	22	47
29	1	1	1	1	1	18	76
30	1	1	1	0	1	32	38
31	0	0	0	0	1	20	19
32	0	1	0	0	1	30	10
33	0	1	1	0	1	18	19
34	0	0	0	0	1	24	24
35	1	1	1	1	1	30	31
36	1	0	1	0	1	35	14
37	0	0	0	0	1	27	11
38	0	1	1	1	1	20	67
39	0	1	0	1	1	22	41
40	0	0	0	0	1	28	7
41	0	0	0	0	1	23	22
42	1	0	0	0	1	40	13
43	1	1	1	1	1	33	46
44	1	1	0	1	1	21	36
45	1	1	1	1	1	35	38
46	0	0	0	0	1	25	7
47	1	1	1	1	1	26	36
48	0	0	0	0	1	25	11
49	1	1	1	1	1	22	151
50	0	0	0	0	1	32	22
51	1	1	1	1	1	25	41
52	0	0	0	1	1	35	32
53	1	1	1	1	1	21	56
54	1	0	0	0	1	41	24
55	0	1	0	0	1	32	16
56	0	1	1	1	1	26	22
57	0	0	0	0	1	21	25
58	0	0	0	0	1	36	13
59	0	0	0	0	1	37	12

6. Exemple/ « épilepsie » /modèle

π_{ij} : probability that individual i has more has 5 (or more) seizures in the 2w period prior visit j

$$\text{logit}\left(\pi_{ij}\right) = \beta_0 + \beta_1 treat_i + \beta_2 age_j + \beta_3 base_i + u_i$$

Visit: not influential; no interaction between covariates

treat (0=Placebo; 1=progabide);

age=(age of patient-30)/10

base=ln(count/4)-ln(ref/4); ref=25

count: seizure count prior to randomisation (8 wks)

$u_i \sim_{\text{iid}} N\left(0, \sigma_u^2\right)$ individual random effect

6. Exemple « epilepsie »/résultats

Occurrence of seizures in epiletics*: marginal vs mixed model analysis

	GLM ML		GEE R=EXCH		Mixed Model ML		Bayes non informatif Posterior mean	
Var.	est±SE	Pvalue	est±SE	Pvalue	est±SE	Pvalue	est±SE	Pvalue
Intercept	0.784±.294		.784±.312		.963±.424		1.046±.492	
Treatment	-.959±.374	0.010	-.959±.428	0.025	-1.132±.537	0.040	-.1.223±.614	0.046
Age	.445±.296	0.133	.445±.327	0.174	.571±.433	0.193	.633±.493	0.200
Base	2.895±.367	<0.0001	2.895±.451	<0.0001	3.397±.561	<0.0001	3.645±.678	<0.0001
Var					1.210±.826		1.806±.1.430	
(Corr)			(.124)		(.269)		(.354)	
-2L	202.2				197.1			
AIC (DIC)	210.2 (136)				207.1		(127)	

*Data from Leppick et al (1985) and Thall and Vail (1990)
Response: Binary: 1=Five or more seizures in the 2 week period prior the visit; 0=otherwise
Intercept=Treatment (Placebo)+Age (30)+log(count ref/4) ; Age=(age in yrs-30)/10; Base=(log(Count/4)-log(ref/4)); ref=25 ; Link=logistic
GLM : Generalized Linear Model , GEE and ML mixed with SAS® Genmod and Nlmixed

6. Exemple « epilepsie »/résultats

Comparaison estimations : m. marginal vs aléatoire

Paramètre	Marginal	Aléatoire	Ratio A/M
Intercept	0.784	0.963	1.23
Traitement	-0.959	-1.132	1.18
Age	0.445	0.571	1.28
Base	2.895	3.397	1.17

$$\beta^A / \beta^M \approx \sqrt{1+\left(\sigma^2 / c^2\right)} \text{ où } c = {}^{15}\!/_{16}\, \pi / \sqrt{3} = 1.7$$

Ici $\sigma^2 = 1.21$ d'où $\beta^A / \beta^M = \boxed{1.19}$

6. Exemple «epilepsie»/genmod GEE

```
proc genmod data=epi descending;
  title epileptics treat base age phi=1 gee exch;
  class indiv ;
  model  seizu=treat base dage/dist=bin type3 wald aggregate=indiv;
  repeated subject=indiv/type=exch corrw;
run;
```

6. Exemple «epilepsie»/genmod GEE

Working Correlation Matrix

	Col1	Col2	Col3	Col4
Row1	1.0000	0.1237	0.1237	0.1237
Row2	0.1237	1.0000	0.1237	0.1237
Row3	0.1237	0.1237	1.0000	0.1237
Row4	0.1237	0.1237	0.1237	1.0000

Analysis Of GEE Parameter Estimates

Empirical Standard Error Estimates

Parameter	Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
Intercept	0.7836	0.3116	0.1728	1.3944	2.51	0.0119
Treat	-0.9587	0.4288	-1.7991	-0.1184	-2.24	0.0254
base	2.8949	0.4512	2.0107	3.7792	6.42	<.0001
dage	0.4449	0.3271	-0.1961	1.0860	1.36	0.1737

6. Exemple «epilepsie»/nlmixed

```
proc  nlmixed data=epi ;
      title epileptics  treat age  base  ml nlmixed;
      parms beta0=0 betaT=0 betaB=0 betaA=0 d=1;
      teta=beta0+(betaT*treat)+(betaA*dage)+(betaB*base)+u;
      expteta=exp(teta);
      p=expteta/(1+expteta);
      model seizu~binary(p);
      random u~normal (0,d)subject=indiv;
   run;
```

6. Exemple «epilepsie»/nlmixed

NOTE: GCONV convergence criterion satisfied.

Fit Statistics	
-2 Log Likelihood	197.1
AIC (smaller is better)	207.1
AICC (smaller is better)	207.3
BIC (smaller is better)	217.5

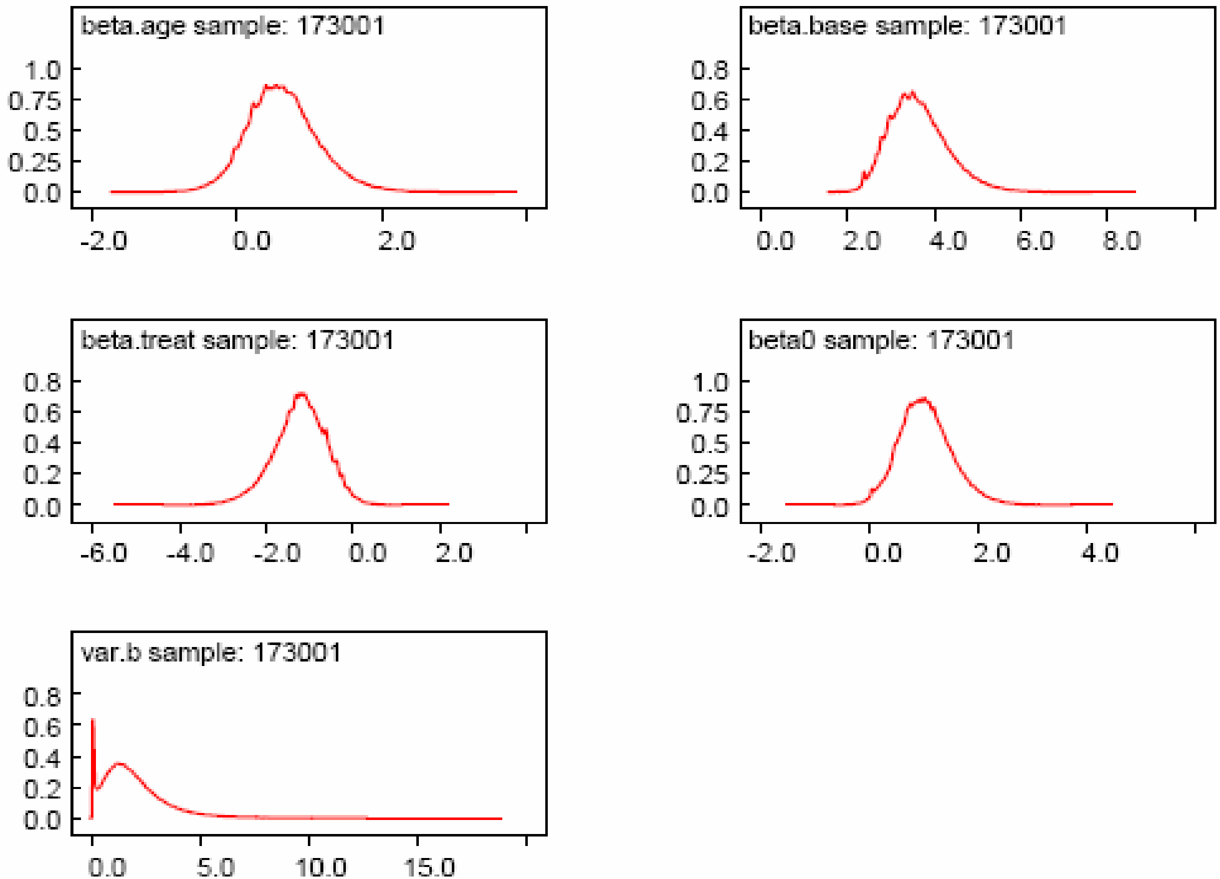
Parameter Estimates

Parameter	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper	Gradient
beta0	0.9628	0.4241	58	2.27	0.0269	0.05	0.1138	1.8118	2.379E-6
betaT	-1.1319	0.5373	58	-2.11	0.0395	0.05	-2.2075	-0.05635	0.00002
betaB	3.3970	0.5612	58	6.05	<.0001	0.05	2.2737	4.5204	8.358E-6
betaA	0.5706	0.4328	58	1.32	0.1925	0.05	-0.2957	1.4369	1.577E-6
d	1.2102	0.8260	58	1.47	0.1483	0.05	-0.4432	2.8637	-4.93E-6

6. Exemple «epilepsie»/winbugs

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
beta.age	0.6327	0.4934	0.005312	-0.2558	0.6025	1.702	38000	173001
beta.base	3.645	0.6779	0.01244	2.522	3.574	5.173	38000	173001
beta.treat	-1.223	0.6136	0.007515	-2.546	-1.194	-0.1128	38000	173001
beta0	1.046	0.4912	0.007028	0.1576	1.015	2.108	38000	173001
var.b	1.806	1.43	0.0297	5.746E-4	1.548	5.35	38000	173001

6. Exemple «epilepsie»/winbugs



6. Exemple « epilepsie »/autre analyse

Occurrence of seizures in epiletics*: marginal vs mixed model analysis

	GLM ML		GEE R=EXCH		Mixed Model ML	
Var.	est±SE	Pvalue	est±SE	Pvalue	est±SE	Pvalue
Intercept	1.140±.339		1.140±.391		1.266±.461	
Treatment	-.899±.377	0.017	-.899±.448	0.045	-1.010±.427	0.060
Age	.227±.149	0.128	.227±.163	0.164	.305±.218	0.167
Pcount	2.967±.386	<0.0001	2.967±.493	<0.0001	3.431±.575	<0.0001
Pcount*2	-.435±.183	0.018	-.435±.118	0.0002	-.500±.225	0.031
Var(corr)			(.129)		1.140 (.266)	
-2L	199.8				195.4	
AIC	209.8				207.4	

*Data from Leppick et al (1985) and Thall and Vail (1990)
Intercept=Treatment (Placebo)+Age (28)+Count (31) ; Age=(age in yrs-28)/5; Pcount=(Count-31)/25; Link=logistic
GLM : Generalized Linear Model , GEE and ML mixed with SAS® Genmod and Nlmixed

7. Compléments

- Modèle mixte marginalisé
  - Exprimer E et V
  - Inférence par quasi-vraisemblance marginale
  - Zeger et al, 1988; Gilmour et al, 1985
  - Breslow & Clayton, 1993
- Données ordinales
  - Modèle à variables latentes: Liu & Agresti, 2005
- Analyse exacte: approche conditionnelle
  - Rasch, 1961; Andersen, 1980
- Validation
  - Outils de diagnostique: Collett, 2003; Fahrmeir & Tutz, 1994
  - Analyse robuste: Cantoni & Ronchetti, 2001

7. Bibliographie

- Collett D., 2003. Modelling binary data, 2nd edition, Chapman & Hall/CRC
- Dobson A.J., 2001. An introduction to GLM, 2nd edition, Chapman & Hall/CRC
- Fahrmeir L., Tutz G.,1994. Multivariate statistical modelling based on GLM, Springer-Verlag