MODELISATION D'UNE REPONSE POLYTOMIQUE

- Cas où les prédicteurs sont qualitatifs: Procédure CATMOD (SAS)

- Cas où les prédicteurs sont quantitatifs: Procédure LOGISTIC (SAS), option glogit

MODELISATION D'UNE REPONSE POLYTOMIQUE

- Présentation des techniques disponibles pour une variable qualitative Y non ordinale

- Présentation des techniques disponibles pour une variable qualitative Y ordinale quand le modèle à rapport de cotes proportionnel ne peut être utilisé.

APPLICATION AU TEST DE L'EFFICACITÉ DU DIFFUSEUR D'IODE RHODIFUSE

1. Les données

- Y = Niveau de goitre = $\begin{cases} 1 = 0 \\ 2 = Ia \\ 3 = Ib \\ 4 = II \end{cases}$

- X_1 = Village = $\begin{cases} 1 = Sirablo \text{ (témoin)} \\ 2 = Woloni \\ 3 = N'Djiba \\ 4 = Sebabougou \end{cases}$

- X_2 = Sexe $\begin{matrix} 1 = Homme \\ 2 = Femme \end{matrix}$

- X_3 = Jour $\begin{matrix} 0 = 0\ j \\ 1 = 180\ j \\ 2 = 360\ j \end{matrix}$

- X_4 = Iode $\begin{matrix} 1 = absence \\ 2 = présence \end{matrix}$

2. Le problème

Étudier la liaison entre la variable dépendante $\boldsymbol{Y}$ = Niveau de goitre

et les variables indépendantes :

$$X_1 = Village$$
$$X_2 = Sexe$$
$$X_3 = Jour$$
$$X_4 = Iode$$

⟺

Étudier la loi de probabilité :

$$\mathbf{Prob}\left(Y = y \mid X_1 = x_1, X_2 = x_2, X_3 = x_3, X_4 = x_4\right)$$

de $\boldsymbol{Y}$ conditionnellement à des valeurs prises par les X_j .

3. Hypothèses à tester

- H_1 : La répartition des niveaux de goitre au temps **O** ne dépend pas du village.

- H_2 : La répartition des niveaux de goitre dépend du sexe : la situation est plus grave chez les femmes que chez les hommes.

- H_3 : La répartition des goitres dépend des jours et de l'iode : la situation se détériore dans le village témoin SIRABLO au cours du temps, alors qu'elle s'améliore dans les autres villages au cours du temps.

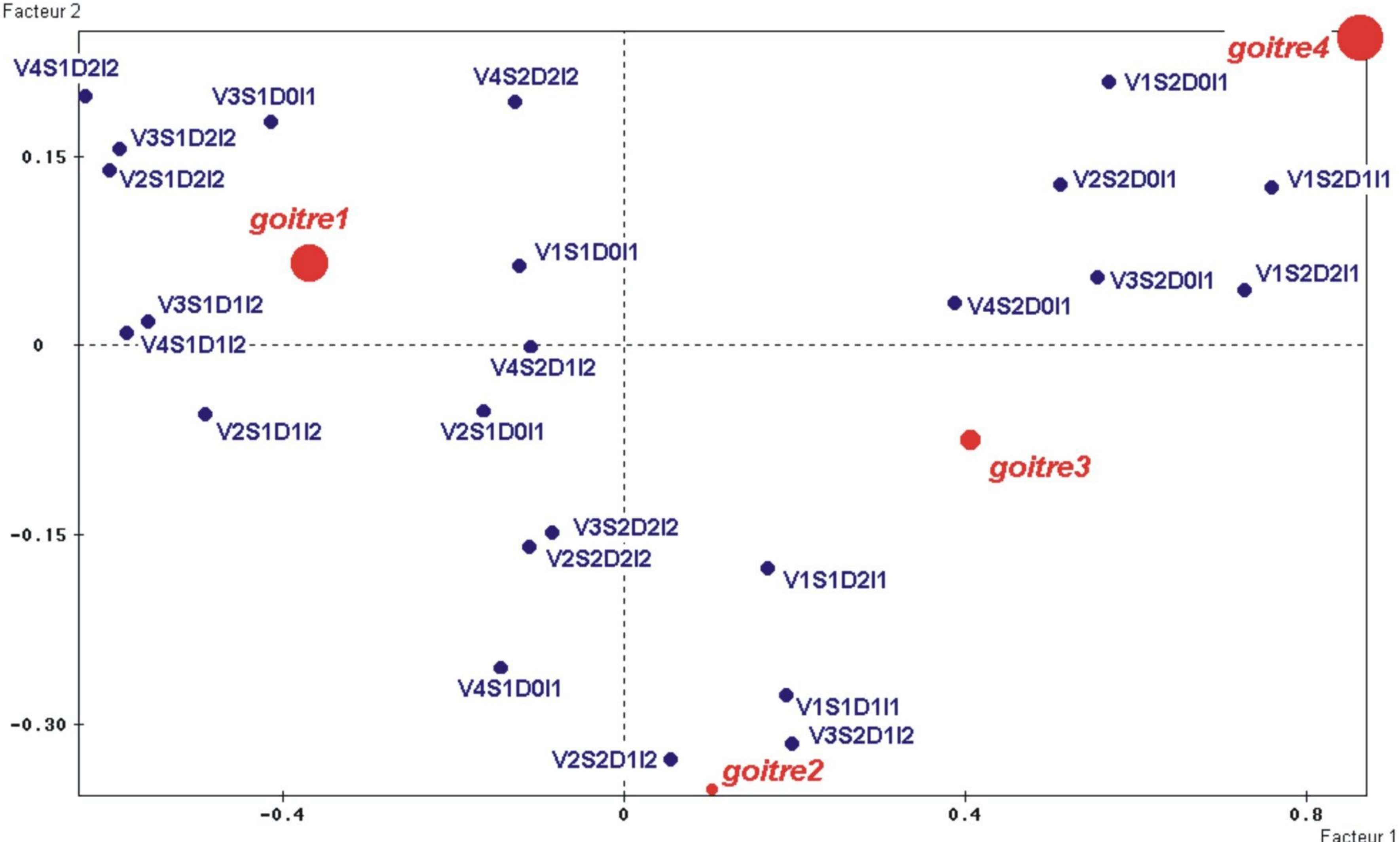
L'ANALYSE STATISTIQUE

I. Analyse des correspondances du tableau des goitres

II. Typologie des profils-lignes

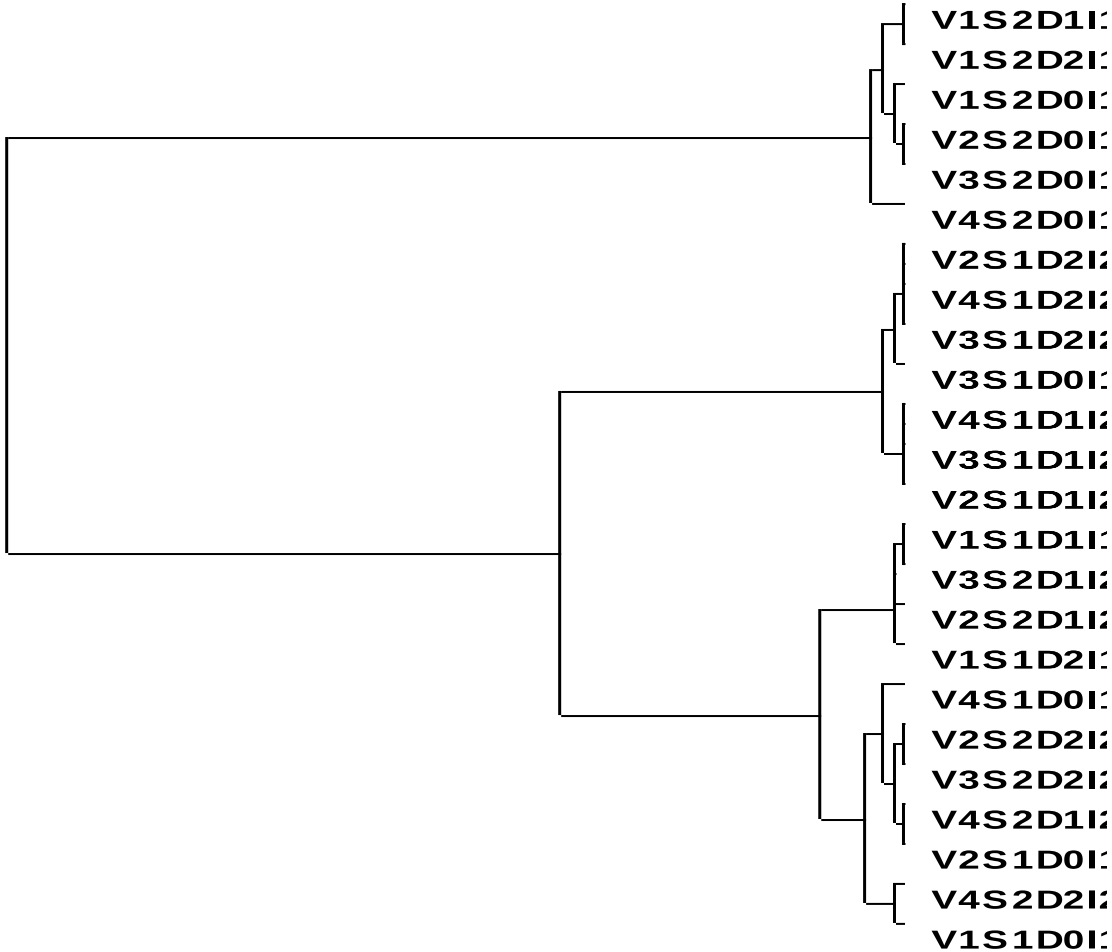
III. Modélisation de la loi de probabilité des goitres en fonction des variables : ***Village***, ***Sexe***, ***Jour***, ***Iode*** à l'aide du modèle multinomial.

I. Analyse des correspondances du tableau des goitres



II. Typologie des profils - lignes

Classification hierarchique directe



III. Utilisation du modèle linéaire généralisé

Les s croisements disponibles $(x_{i1}, \dots x_{ik})$ des variables $X_1 \dots X_k$ définissent s populations. On note Π_i la loi de probabilité de Y sur la population i.

On cherche à relier linéairement q fonctions de réponse :

$F_h(\Pi_i)$ aux caractéristiques de la population i

$$F_h(\Pi_i) = x_i' \beta_h$$

où x_i =vecteur ligne construit à partir des caractéristiques de la population i

et $\beta_h =$ vecteur colonne de paramètres

Exemples de fonction de réponse

$$Log\ {}^{P_1}\!/_{P_4} = \alpha_1 + x'\beta_1$$

$$Log\ {}^{P_2}\!/_{P_4} = \alpha_2 + x'\beta_2$$

$$Log\ {}^{P_3}\!/_{P_4} = \alpha_3 + x'\beta_3$$

APPLICATION À L'EXEMPLE DES GOITRES

1. Choix de la fonction de réponse

Identité

$$F_h(\Pi_i) = \Pi_{ih} = \text{Prob}(Y = h \mid i) \qquad h = 1 \text{ à } 3$$

Logit généralisé

$$F_h(\Pi_i) = Log\left(\Pi_{ih} / \Pi_{i4}\right) \qquad h = 1 \text{ à } 3$$

(généralisation de la régression logistique)

Logit adjacent

$$F_h\left(\Pi_i\right) = \mathrm{Log}\left(\Pi_{i(h+1)} / \Pi_{ih}\right) \qquad h = 1 \text{ à } 3$$

Logit cumulé

$$\boldsymbol{F_h\left(\Pi_i\right) = Log\left[\frac{\mathbf{Prob}\left(Y > h \mid i\right)}{\mathbf{Prob}\left(Y \leq h \mid i\right)}\right] \qquad h = 1 \text{ à } 3}$$

Moyenne

$$\boldsymbol{F\left(\Pi_i\right) = \sum_{j=1}^{4} j\,\Pi_{ij}}$$

Remarque

Pour $\boldsymbol{F_h(\Pi_i)}$ SAS donne les r-1 premières valeurs (ici les 3 premières valeurs). (La somme de toutes les probabilités est égale à **1**).

Parmi ces cinq options l'option identité et l'option logit généralisé sont utilisables pour toutes les données nominales. En revanche logit adjacent, logit cumulé et moyenne ne doivent être utilisées que pour une variable Y ordinale.

2. Choix du modèle

Pour $\boldsymbol{h = 1\ \grave{a}\ 3}$

$\boldsymbol{F_h(\Pi_i) =}$ Village, Sexe, Village*Sexe,

Iode (Village = 2), Iode (Village = 3),

Iode (Village = 4), Sexe * Iode

Jour (Iode = Abs), Jour (Iode = Prés)

Les problèmes

1. Le modèle probabiliste

2. Estimation des paramètres du modèle

   - GLS en général (moindres carrés généralisés)

   - Maximum de vraisemblance pour le logit généralisé

3. Test sur les variables explicatives

4. Test d’adéquation du modèle

5. Étude de contrastes

Remarques

- Avantages du maximum de vraisemblance :
  → permet de tester plus facilement les cases vides.

- Dans la pratique les « **0** » gênent avec la technique GLS
  (option ADDCELL dans SAS)

3. Le modèle probabiliste

Population	Réponse Y: $1 \dots j \dots r$	
1	$n_{11} \quad \dots \quad n_{1j} \quad \dots \quad n_{1r}$	
i	$n_{i1} \quad \dots \quad n_{ij} \quad \dots \quad n_{ir}$	$\rightarrow \sum_{j=1}^{r} n_{ij} = n_i$
s	$n_{s1} \quad \dots \quad n_{sj} \quad \dots \quad n_{sr}$	

1		
i	$\boxed{P_{i1} \quad \dots \quad P_{ij} \quad \dots \quad P_{ir}}$	P_i
s		

$$\Pi_{ij} = \text{Prob}\left(Y = j \text{ dans la population } i\right)$$

$$P_{ij} = \frac{n_{ij}}{n_i} = \text{estimation de } \Pi_{ij}$$

Loi multinomiale

$$\text{Prob}\,\underbrace{\left(P_{i1},\ldots,P_{ir}\right)}_{P_i} = n_i\,!\,\frac{\Pi_{i1}^{n_{i1}}\ldots\Pi_{ir}^{n_{ir}}}{\mathrm{n_{i1}!\ldots n_{ir}!}}$$

$$E\left(P_i\right) = \Pi_i = \left(\Pi_{i1},\ldots,\Pi_{ir}\right)$$

$$V\left(P_i\right) = \begin{bmatrix} \frac{\Pi_{i1}\left(1-\Pi_{i1}\right)}{n_i} & -\frac{\Pi_{i1}\,\Pi_{i2}}{n_i} & \cdots & -\frac{\Pi_{i1}\,\Pi_{ir}}{n_i} \\ & \frac{\Pi_{i2}\left(1-\Pi_{i2}\right)}{n_i} & \cdots & -\frac{\Pi_{i2}\,\Pi_{ir}}{n_i} \\ & & \ddots & \frac{\Pi_{\mathrm{ir}}\left(1-\Pi_{\mathrm{ir}}\right)}{\mathrm{n_i}} \end{bmatrix}$$

$$\mathbf{Prob}\left(\underbrace{P_1, \ldots, P_s}_{P}\right) = \prod_{i=1}^{s} \mathbf{Prob}\left(P_i\right)$$

$$E(P) = \Pi = \left(\Pi_1, \ldots, \Pi_s\right)$$

$$V(P) = \begin{bmatrix} V(P_1) & & 0 \\ & \ddots & \\ 0 & & V(P_s) \end{bmatrix}$$

matrice diagonale par blocs

estimée par V en remplaçant les Π_{ij} par P_{ij}

4. Modélisation des $\prod_{ij}$

$$F_h(\Pi_i) = \begin{cases} \Pi_{ih} = \mathrm{Prob}\left(Y = h \mid \mathrm{Pop}\ i\right) & \textbf{Identité} \\ \mathrm{Log}\left(\Pi_{ih} / \Pi_{ir}\right) & \textbf{Logit généralisé} \\ \mathrm{Log}\left[\dfrac{\mathrm{Prob}\left(Y > h \mid \mathrm{Pop}\ i\right)}{\mathrm{Prob}\left(Y \leq h \mid \mathrm{Pop}\ i\right)}\right] & \textbf{Logit cumulé} \end{cases}$$

$$F_h(\Pi_i) = \beta_0^h + \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \end{array}\begin{bmatrix} \beta_1^h \\ \beta_2^h \\ \beta_3^h \\ -\beta_1^h - \beta_2^h - \beta_3^h \end{bmatrix} + \begin{array}{c} H \\ F \end{array}\begin{bmatrix} \beta_4^h \\ -\beta_4^h \end{bmatrix} + \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \end{array}\left[\begin{array}{c|c} \beta_5^h & -\beta_5^h \\ \beta_6^h & -\beta_6^h \\ \beta_7^h & -\beta_7^h \\ -\beta_5^h - \beta_6^h - \beta_7^h & \beta_5^h + \beta_6^h + \beta_7^h \end{array}\right]$$

Village **Sexe** **Village * Sexe**

$$+ \begin{array}{c} \\ 2 \\ 3 \\ 4 \end{array}\begin{array}{c} \begin{array}{cc} 1 & 2 \end{array} \\ \left[\begin{array}{c|c} \beta_8^h & -\beta_8^h \\ \beta_9^h & -\beta_9^h \\ \beta_{10}^h & -\beta_{10}^h \end{array}\right] \end{array} + \begin{array}{c} \\ H \\ F \end{array}\begin{array}{c} \begin{array}{cc} 1 & 2 \end{array} \\ \left[\begin{array}{c|c} \beta_{11}^h & -\beta_{11}^h \\ -\beta_{11}^h & \beta_{11}^h \end{array}\right] \end{array} + \begin{array}{c} \\ 1 \\ 2 \end{array}\begin{array}{c} \begin{array}{ccc} 0 & 180 & 360 \end{array} \\ \left[\begin{array}{c|c|c} \beta_{12}^h & \beta_{13}^h & -\beta_{12}^h - \beta_{13}^h \\ \hline & \beta_{14}^h & -\beta_{14}^h \end{array}\right] \end{array}$$

Iode (Village 2, 3, 4) **Sexe * Iode** **Jour (Iode)**

Design Matrix

Échantillon	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	1	1	0	0	1	1	0	0	0	0	0	1	1	0	0	0
2	1	1	0	0	1	1	0	0	0	0	0	1	0	1	0	0
3	1	1	0	0	1	1	0	0	0	0	0	1	-1	-1	0	0
4	1	1	0	0	-1	-1	0	0	0	0	0	-1	1	0	0	0
5	1	1	0	0	-1	-1	0	0	0	0	0	-1	0	1	0	0
6	1	1	0	0	-1	-1	0	0	0	0	0	-1	-1	-1	0	0
7	1	0	1	0	1	0	1	0	1	0	0	1	1	0	0	0
8	1	0	1	0	1	0	1	0	-1	0	0	-1	0	0	0	1
9	1	0	1	0	1	0	1	0	-1	0	0	-1	0	0	-1	-1
10	1	0	1	0	-1	0	-1	0	1	0	0	-1	1	0	0	0
11	1	0	1	0	-1	0	-1	0	-1	0	0	1	0	0	0	1
12	1	0	1	0	-1	0	-1	0	-1	0	0	1	0	0	-1	-1
13	1	0	0	1	1	0	0	1	0	1	0	1	1	0	0	0
14	1	0	0	1	1	0	0	1	0	-1	0	-1	0	0	0	1
15	1	0	0	1	1	0	0	1	0	-1	0	-1	0	0	-1	-1
16	1	0	0	1	-1	0	0	-1	0	1	0	-1	1	0	0	0
17	1	0	0	1	-1	0	0	-1	0	-1	0	1	0	0	0	1
18	1	0	0	1	-1	0	0	-1	0	-1	0	1	0	0	-1	-1
19	1	-1	-1	-1	1	-1	-1	-1	0	0	1	1	1	0	0	0
20	1	-1	-1	-1	1	-1	-1	-1	0	0	-1	-1	0	0	0	1
21	1	-1	-1	-1	1	-1	-1	-1	0	0	-1	-1	0	0	-1	-1
22	1	-1	-1	-1	-1	1	1	1	0	0	1	-1	1	0	0	0
23	1	-1	-1	-1	-1	1	1	1	0	0	-1	1	0	0	0	1
24	1	-1	-1	-1	-1	1	1	1	0	0	-1	1	0	0	-1	-1

5. Estimation des Coefficients

La méthode GLS (generalized least squares) prend en compte le fait que les p_{ij} ont des variances inégales et sont corrélées entre elles. On minimise :

$$\boxed{SCR = \left(F - \hat{F}\right)' \, \widehat{Var\left(F\right)}^{-1} \left(F - \hat{F}\right)}$$

où : • $F = \left(F_h\left(p_i\right)\right)$

•• $\hat{F} = \left(\widehat{F_h\left(\Pi_i\right)}\right)$ = vecteur des $F_h\left(\Pi_i\right)$ estimées à l'aide du modèle étudié

••• $\widehat{Var\left(F\right)} =$ estimation de la matrice de variances/ covariances de F

Analysis of Weighted Least Squares Estimates

Parameter		Nombre de fonctions	Estimation	Erreur std	Khi-2	Pr > Khi 2
Intercept		1	0.4735	0.0110	1867.62	<.0001
		2	0.1679	0.00839	400.04	<.0001
		3	0.2286	0.0101	514.59	<.0001
village	1	1	-0.1069	0.0146	53.20	<.0001
	1	2	-0.0347	0.00951	13.34	0.0003
	1	3	0.0798	0.0133	35.89	<.0001
	2	1	0.0177	0.0128	1.93	0.1653
	2	2	0.0122	0.00901	1.82	0.1773
	2	3	-0.00601	0.0111	0.29	0.5893
	3	1	0.0364	0.0142	6.56	0.0104
	3	2	-0.00249	0.00997	0.06	0.8029
	3	3	-0.00994	0.0124	0.64	0.4223
sexe	1	1	0.1343	0.00705	362.79	<.0001
	1	2	-0.0126	0.00505	6.19	0.0128
	1	3	-0.0557	0.00612	82.79	<.0001
village*sexe	1 1	1	-0.0269	0.0149	3.25	0.0714
	1 1	2	-0.00131	0.0105	0.02	0.9012
	1 1	3	0.0368	0.0137	7.20	0.0073
	2 1	1	0.00394	0.0122	0.10	0.7461
	2 1	2	-0.00509	0.00885	0.33	0.5653
	2 1	3	-0.00568	0.0104	0.30	0.5834
	3 1	1	0.0548	0.0135	16.54	<.0001
	3 1	2	-0.0246	0.00969	6.42	0.0113
	3 1	3	-0.0350	0.0117	9.03	0.0027
iode(village=2)	1 1	1	-0.1128	0.0178	39.94	<.0001
	1 1	2	-0.0140	0.0127	1.20	0.2732
	1 1	3	0.0639	0.0158	16.44	<.0001
iode(village=3)	1 1	1	-0.0767	0.0199	14.92	0.0001
	1 1	2	-0.0117	0.0138	0.72	0.3976
	1 1	3	0.0233	0.0171	1.85	0.1739
iode(village=4)	1 1	1	-0.1561	0.0175	79.85	<.0001
	1 1	2	0.0601	0.0133	20.31	<.0001
	1 1	3	0.0340	0.0144	5.56	0.0184

sexe*iode	1 1	1	-0.00094	0.00860	0.01	0.9132
	1 1	2	0.0173	0.00632	7.48	0.0062
	1 1	3	0.0111	0.00715	2.41	0.1205
jour(iode=1)	0 1	1	0.0898	0.0189	22.56	<.0001
	0 1	2	-0.0492	0.0127	14.92	0.0001
	0 1	3	-0.0210	0.0183	1.31	0.2517
	1 1	1	-0.0466	0.0189	6.07	0.0137
	1 1	2	0.0312	0.0150	4.31	0.0380
	1 1	3	-0.00187	0.0192	0.01	0.9223
jour(iode=2)	0 1	1	-0.0763	0.0187	16.73	<.0001
	0 1	2	0.0681	0.0141	23.35	<.0001
	0 1	3	0.00931	0.0138	0.46	0.4998
	1 1	1	.	.	.	.
	1 1	2	.	.	.	.
	1 1	3	.	.	.	.

Exemple

$$\hat{\Pi}_{i1} = \widehat{\mathbf{Prob}}\left(Y = 1 \mid \mathbf{Pop}\ i\right)$$

$$= 0{,}47 + \begin{array}{c|c} & \text{} \\ \begin{array}{c}1\\2\\3\\4\end{array} & \begin{bmatrix}-\cdot 11\\ \cdot 02\\ \cdot 03\\ \cdot 06\end{bmatrix}\end{array} + \begin{array}{c}H\\F\end{array}\begin{bmatrix}\cdot 13\\ -\cdot 13\end{bmatrix} + \begin{array}{cc} \begin{array}{c}1\\2\\3\\4\end{array} & \left[\begin{array}{c|c}-\cdot 03 & \cdot 03\\ 0 & 0\\ \cdot 05 & -\cdot 05\\ -\cdot 02 & \cdot 02\end{array}\right] \\ & \begin{array}{cc}\quad H & \quad\quad F\end{array}\end{array}$$

Village Sexe Village * Sexe

$$+ \begin{array}{cc} \begin{array}{c}2\\3\\4\end{array} & \begin{bmatrix}-\cdot 11 & \cdot 11\\ -\cdot 08 & \cdot 08\\ -\cdot 16 & \cdot 16\end{bmatrix} \\ & \begin{array}{cc}1 & \quad\quad 2\end{array}\end{array} + \begin{array}{cc} \begin{array}{c}H\\F\end{array} & \begin{bmatrix}0 & 0\\ 0 & 0\end{bmatrix} \\ & \begin{array}{cc}1 & 2\end{array}\end{array} + \begin{array}{cc} \begin{array}{c}1\\2\end{array} & \left[\begin{array}{c|c|c}\cdot 09 & -\cdot 05 & -\cdot 04\\ & -\cdot 08 & \cdot 08\end{array}\right] \\ & \begin{array}{ccc}0 & \quad 180 & \quad 360\end{array}\end{array}$$

Iode (Village 2, 3, 4) Sexe*Iode Jour (Iode)

Exemple de calcul d'une probabilité

$$\hat{\Pi}_{11} = 0{,}47 - \cdot 11 + \cdot 13 - \cdot 03 + \cdot 09 = 0 \cdot 56$$

à comparer à $\boldsymbol{p}_{11} = \dfrac{106+1}{175+4} = 0{,}5977$

Predicted Values for Response Functions

village	sexe	iode	jour	Nb fcn	-----Observé(s)----- Fonction	Erreur std	------Prédit(e)----- Fonction	Erreur std	Résidus
1	1	1	0	1	0.597765	0.03665	0.562911	0.016938	0.034854
				2	0.072626	0.019398	0.087365	0.016938	-0.01474
				3	0.26257	0.032889	0.279571	0.016938	-0.017
1	1	1	1	1	0.391026	0.03907	0.426496	0.020966	-0.03547
				2	0.205128	0.032329	0.167716	0.020966	0.037412
				3	0.301282	0.036735	0.298699	0.020966	0.002583
1	1	1	2	1	0.419355	0.039635	0.429926	0.020579	-0.01057
				2	0.154839	0.029057	0.154584	0.020579	0.000254
				3	0.329032	0.03774	0.323445	0.020579	0.005587
......									
4	2	1	0	1	0.375	0.031784	0.358417	0.022377	0.016583
				2	0.176724	0.025042	0.16822	0.022377	0.008504
				3	0.206897	0.026595	0.218453	0.022377	-0.01156
4	2	2	1	1	0.582524	0.034359	0.578964	0.014226	0.00356
				2	0.131068	0.023513	0.131767	0.014226	-0.0007
				3	0.194175	0.02756	0.193663	0.014226	0.000512
4	2	2	2	1	0.630332	0.033231	0.65527	0.014288	-0.02494
				2	0.061611	0.016553	0.063625	0.014288	-0.00201
				3	0.199052	0.027488	0.184349	0.014288	0.014703

6. Tests sur les variables explicatives

Exemple VILLAGE

Test $H_0 \;:\; \beta_1^h = \beta_2^h = \beta_3^h = 0$ pour $h = 1, \ldots 3$

Statistique de Wald

$$Q = \hat{\beta}'_{\text{village}} \; Var\left(\hat{\beta}_{\text{village}}\right)^{-1} \; \hat{\beta}_{\text{village}} \qquad \text{sous } H_0 \;:\; Q \text{ suit } \chi^2_9$$

Analysis of Variance

Source	DF	Chi-Square	Pr > ChiSq
Intercept	3	12968.97	<.0001
village	9	123.91	<.0001
sexe	3	465.13	<.0001
village*sexe	9	32.52	0.0002
iode(village=2)	3	65.97	<.0001
iode(village=3)	3	35.41	<.0001
iode(village=4)	3	90.12	<.0001
sexe*iode	3	28.66	<.0001
jour(iode=1)	6	30.27	<.0001
jour(iode=2)	3*	25.41	<.0001
Residual	27	16.45	0.9438

NOTE: Effects marked with '*' contain one or more redundant or restricted parameters.

Remarque

$\boldsymbol{Q}$ = SCR (Modèle sans village) - SCR (modèle complet)

Analogie avec les sommes de carrés de type III dans GLM.

Résultat

$$\boldsymbol{Q = 124 \qquad P.Value = 0{,}0000}$$

La variable village a un apport significatif dans le modèle.

7. Tests d'adéquation du modèle

Test H_0 : Modèle étudié exact

Statistique $SCR = (F - \hat{F})' \widehat{Var}(F)^{-1} (F - \hat{F})$

Sous H_0 :

$$SCR = \underbrace{F' Var(F)^{-1} F}_{\chi^2_n} - \underbrace{\hat{F}' \widehat{Var}(F)^{-1} \hat{F}}_{\chi^2_d}$$

Nombre de fonctions de réponse $n = 3 \times 24 = 72$

Nombre de paramètres du modèle $d = 3 \times 15 = 45$

SCR suit $\chi^2_{\underbrace{n-d}_{27}}$

Exemple

$$\boldsymbol{SCR = 16{,}45}$$

Niveau de signification **0,94**

On ne rejette pas $\boldsymbol{H_0}$.

8. Modélisation de type logit généralisé

$$F_h(\Pi_i) = Log\left(\Pi_{ih}/\Pi_{ir}\right)$$

Du modèle écrit précédemment on déduit :

$$\begin{cases} \Pi_{ih} = \dfrac{e^{x_i\,\beta_h}}{1+\sum\limits_{h=1}^{r-1} e^{x_i\,\beta_h}} \quad ; \quad h=1\ \ldots\ r-1 \\[2em] \Pi_{ir} = \dfrac{1}{1+\sum\limits_{h=1}^{r-1} e^{x_i\,\beta_h}} \end{cases}$$

avec $x_i =$ caractéristiques du profil i

$\beta^h =$ vecteur des β_j^h

Estimation des β_j^h par le maximum de vraisemblance

Multinomiale

$$\text{Prob}\left(\ldots p_{ih}\ldots\right)=\prod_{i=1}^{s} n_i!\,\frac{\Pi_{i1}^{n_{i1}}\cdots\Pi_{ir}^{n_{ir}}}{n_{i1}!\cdots n_{ir}!}$$

Modèle

$$\Pi_{ih}=f_k\underbrace{\left(\cdots\beta_j^l\cdots\right)}_{\beta}$$

Vraisemblance

$$L\left(\text{Modèle}\right)=\prod_{i=1}^{s} n_i!\,\frac{\left[f_1(\beta)\right]^{n_{i1}}\cdots\left[f_r(\beta)\right]^{n_{ir}}}{n_{i1}!\cdots n_{ir}!}$$

On recherche β maximisant la vraisemblance.

Le modèle saturé

Si le nombre de paramètres indépendants β_j^k est égal au nombre de probabilités indépendantes Π_{ih} le modèle est dit **saturé**.

Alors :

$$\hat{\Pi}_{ih} = \frac{n_{ih}}{n_i} = p_{ih}$$

L (modèle saturé)

$$= \prod_{i=1}^{s} n_i! \frac{p_{i1}^{n_{i1}} \cdots p_{ir}^{n_{ir}}}{n_{i1}! \cdots n_{ir}!}$$

Test d'ajustement

Test H_0 **:** Modèle étudié exact

Statistique utilisée

$$D = -2\,\mathrm{Log}\left[\frac{L(\text{Modèle étudié})}{L(\text{Modèle saturé})}\right]$$

(s'interprète comme une somme résiduelle :

si $D = 0$ modèle étudié $=$ modèle saturé)

Sous H_0

$D \rightsquigarrow \chi^2_{n-d}$

n = nombre de fonctions de réponse
d = nombre de paramètres du modèle

Tests sur les variables explicatives

Exemple Village

Test $$Ho\ :\ \beta_1^h = \beta_2^h = \beta_3^h = 0 \quad h = 1, \cdots 3$$

Statistique $$D = -2\ \text{Log}\left[\frac{L(\text{Modèle sans village})}{L(\text{Modèle complet})}\right]$$

Sous $Ho : D \leadsto \chi_9^2$

Maximum Likelihood Analysis of Variance

Source	DF	Chi-Square	Pr > ChiSq
Intercept	3	371.22	<.0001
village	9	123.07	<.0001
sexe	3	291.31	<.0001
village*sexe	9	21.94	0.0091
iode(village=2)	3	67.56	<.0001
iode(village=3)	3	36.38	<.0001
iode(village=4)	3	84.76	<.0001
sexe*iode	3	8.24	0.0413
jour(iode=1)	6	30.46	<.0001
jour(iode=2)	3*	24.14	<.0001
Likelihood Ratio	27	19.62	0.8463

NOTE: Effects marked with '*' contain one or more
redundant or restricted parameters.

9. Étude du modèle moyenne

$$F\left(\Pi_i\right)=\sum_{j=1}^{4} j\,\Pi_{ij}$$

$$=\beta_0+\begin{array}{c}1\\2\\3\\4\end{array}\begin{bmatrix}\beta_1\\ \beta_2\\ \beta_3\\ -\beta_1-\beta_2-\beta_3\end{bmatrix}+\begin{array}{c}H\\F\end{array}\begin{bmatrix}\beta_4\\ -\beta_4\end{bmatrix}+\begin{array}{c}1\\2\\3\\4\end{array}\left[\begin{array}{c|c}\beta_5 & -\beta_5\\ \beta_6 & -\beta_6\\ \beta_7 & -\beta_7\\ -\beta_5-\beta_6-\beta_7 & \beta_5+\beta_6+\beta_7\end{array}\right]$$

Village Sexe Village * Sexe

$$+\begin{array}{c}2\\3\\4\\ \\ \end{array}\begin{array}{c}\left[\begin{array}{c|c}\beta_8 & -\beta_8\\ \beta_9 & -\beta_9\\ \beta_{10} & -\beta_{10}\end{array}\right]\\ \begin{array}{cc}1 & \quad 2\end{array}\end{array}+\begin{array}{c}H\\F\\ \\ \end{array}\begin{array}{c}\left[\begin{array}{c|c}\beta_{11} & -\beta_{11}\\ -\beta_{11} & \beta_{11}\end{array}\right]\\ \begin{array}{cc}1 & \quad 2\end{array}\end{array}+\begin{array}{c}1\\2\\ \\ \end{array}\begin{array}{c}\left[\begin{array}{c|c|c}\beta_{12} & \beta_{13} & -\beta_{12}-\beta_{13}\\ \times & \beta_{14} & -\beta_{14}\end{array}\right]\\ \begin{array}{ccc}0 & 180 & \quad 360\end{array}\end{array}$$

Iode (Village 2,3,4) Sexe * Iode Jour (Iode)

MODELE FONCTION DE REPONSE: MOYENNE

Analysis of Variance

Source	DF	Chi-Square	Pr > ChiSq
Intercept	1	6603.20	<.0001
village	3	88.43	<.0001
sexe	1	455.33	<.0001
village*sexe	3	7.77	0.0509
iode(village=2)	1	63.58	<.0001
iode(village=3)	1	26.65	<.0001
iode(village=4)	1	69.60	<.0001
sexe*iode	1	5.34	0.0208
jour(iode=1)	2	12.58	0.0019
jour(iode=2)	1*	6.10	0.0135
Residual	9	8.44	0.4910

NOTE: Effects marked with '*' contain one or more redundant or restricted parameters.

Analysis of Weighted Least Squares Estimates

Effect	Parameter	Estimate	Standard Error	Chi-Square	Pr > ChiSq
Intercept		2.0203	0.0249	6603.20	<.0001
village	1	0.3055	0.0333	84.01	<.0001
	2	-0.0661	0.0280	5.55	0.0184
	3	-0.0970	0.0315	9.50	0.0021
sexe	1	-0.3226	0.0151	455.33	<.0001
village*sexe	1 1	0.0433	0.0345	1.58	0.2093
	2 1	0.00200	0.0254	0.01	0.9374
	3 1	-0.0760	0.0284	7.14	0.0075
iode(village=2)	1 1	0.3117	0.0391	63.58	<.0001
iode(village=3)	1 1	0.2266	0.0439	26.65	<.0001
iode(village=4)	1 1	0.3185	0.0382	69.60	<.0001
sexe*iode	1 1	-0.0431	0.0186	5.34	0.0208
jour(iode=1)	0 1	-0.1618	0.0458	12.49	0.0004
	1 1	0.0926	0.0467	3.93	0.0474
jour(iode=2)	0 1	0.0853	0.0345	6.10	0.0135
	1 1	.	.	.	.

Estimation du modèle « Moyenne »

$$F(\Pi_i) = \sum_{j=1}^{4} j\,\Pi_{ij}$$

$$\approx\ 2.02 + \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \end{array}\!\begin{bmatrix} \cdot 31 \\ -\cdot 07 \\ -\cdot 10 \\ -\cdot 14 \end{bmatrix} + \begin{array}{c} H \\ F \end{array}\!\begin{bmatrix} -\cdot 32 \\ +\cdot 32 \end{bmatrix} + \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \end{array}\!\left[\begin{array}{c|c} \cdot 04 & -\cdot 04 \\ \cdot 00 & \cdot 00 \\ -\cdot 08 & \cdot 08 \\ \cdot 04 & -\cdot 04 \end{array}\right]$$

Village Sexe Village * Sexe

$$+ \begin{array}{c} 2 \\ 3 \\ 4 \end{array}\!\left[\begin{array}{c|c} \cdot 31 & -\cdot 31 \\ \cdot 22 & -\cdot 22 \\ \cdot 32 & -\cdot 32 \end{array}\right] + \begin{array}{c} H \\ F \end{array}\!\left[\begin{array}{c|c} -\cdot 04 & \cdot 04 \\ \cdot 04 & -\cdot 04 \end{array}\right] + \begin{array}{c} 1 \\ 2 \end{array}\!\left[\begin{array}{c|c|c} -\cdot 16 & \cdot 09 & \cdot 07 \\ \times & \cdot 09 & -\cdot 09 \end{array}\right]$$

Iode (Village 2,3,4) Sexe * Iode Jour (Iode)

10. Étude des contrastes

Objectif Comparaison de moyenne

$$[F_1(\Pi_i),\cdots,F_r(\Pi_i)] = x_i\beta$$

Test

$$H_0 \qquad L\beta = 0$$

(condition : somme des coefficients nulle)

Statistique de Wald

$$Q = \left(L\hat{\beta}\right)' \left[Var\left(L\hat{\beta}\right)\right]^{-1} L\hat{\beta}$$

Sous H_o

$$Q \rightsquigarrow \chi^2_{[\text{rang de } L]}$$

Test 1 Le niveau moyen des goitres sur toute la durée de l'expérimentation est significativement supérieur dans le village témoin de Sirablo par rapport aux autres villages :

$$\begin{cases} H_0 : \beta_1 = \dfrac{\beta_2 + \beta_3 - \beta_1 - \beta_2 - \beta_3}{3} \\ H_1 : \beta_1 > \dfrac{\beta_2 + \beta_3 - \beta_1 - \beta_2 - \beta_3}{3} \end{cases} \Leftrightarrow \begin{cases} \beta_1 = 0 \\ \beta_1 > 0 \end{cases}$$

Statistique de Wald $$Q = \frac{\hat{\beta}_1^2}{s\left(\hat{\beta}_1\right)^2} = 84$$

Conclusion: On rejette H_0 au profit de $H_1\left(\hat{\beta}_1 = \cdot 31\right)$

Test 2 « Les trois villages où on a installé le ‘Rhodifuse’ Iode sont équivalents. »

$$H_0 \quad : \quad \beta_2 = \beta_3 \quad ; \quad \beta_2 = -\beta_1 - \beta_2 - \beta_3$$

$$(\beta_2 - \beta_3 = 0) \qquad \beta_1 + 2\beta_2 + \beta_3 = 0$$

Statistique de Wald

$$Q = 3 \cdot 17 \quad (N.S = 0{,}2)$$

On ne rejette pas H_0

Test 3 Il y a un effet Iode dans chaque village. Cet effet est un peu moins fort dans le village de N'DJIBA.

$$\begin{cases} H_0 : \beta_9 = \dfrac{\beta_8 + \beta_{10}}{2} \\ H_1 : \beta_9 < \dfrac{\beta_8 + \beta_{10}}{2} \end{cases}$$

$Q = 4{,}47$ (Niveau de signification $= \cdot 034$)

$\Rightarrow$ On rejette H_0

Test 4 Comparaison des villages sans apport d'iode

Test

$$\beta_1 = \beta_2 + \beta_8 = \beta_3 + \beta_9 = -\beta_1 - \beta_2 - \beta_3 + \beta_{10}$$

$$Q = 5{,}42 \ \left(\text{N.S.} = 0{,}1433\right)$$

Conclusion

On ne rejette pas l'hypothèse d'homogénéité des quatre villages, quand il n'y a pas d'apport d'iode.

Analyse de Contrastes

	Contrast	DF	Chi-Square	Pr > ChiSq
TEST 2	**villages test**	**2**	**3.17**	**0.2045**
TEST 3	**iode n_jiba**	**1**	**4.47**	**0.0344**
TEST 4	**villages j0**	**3**	**5.42**	**0.1433**