

Comparaison du Logiciel Glim4 et de la Procédure
GENMOD du Logiciel SAS pour les Modèles
Linéaires Généralisés

Marc CARAYON

août-septembre 1996

Je tiens à remercier M.Saint-Pierre (CICT)
et M.Mathieu (Laboratoire Statistiques et Probabilités)
pour leur aide si précieuse.

Table des matières

Introduction	1
1 Les Modèles Linéaires Généralisés	2
1.1 Définitions	2
1.2 Propriétés	3
1.3 Exemples	4
1.3.1 Modèle Gaussien	4
1.3.2 Modèle Gamma	4
1.3.3 Modèle Poissonnien	5
1.3.4 Modèle Binomial	5
2 Comparaison de Glim4 et de SAS-GENMOD	6
2.1 Brève explication de Glim4 et SAS-GENMOD	6
2.2 Comparaison pour les Modèles Linéaires Généralisés	10
2.2.1 Avant-Propos	10
2.2.2 Intercept et paramètre d'échelle	14
2.2.3 Liens et Distributions	15
2.2.4 Paramètres Estimés et Erreurs Standards	18
2.2.5 Résidus, Valeurs prédites et Erreurs standards	19
2.2.6 Tests et Intervalles de confiance	20
2.2.7 Corrélation et Covariance	21
2.2.8 Poids et offset	22
2.2.9 Autres	22
3 Applications	27
3.1 Modèle Gaussien	27
3.2 Modèle Gamma	29
3.3 Modèle Poissonnien	32
3.4 Modèle Binomial	35
Conclusion	41

Introduction

Ce manuel est destiné à des personnes sachant se servir de l'un de ces deux logiciels (Glim4 ou SAS sur le système UNIX) et désirant apprendre à se servir de l'autre.

SAS est sans doute le logiciel le plus répandu et peut être même le plus complet.

SAS est un logiciel fragmenté en plusieurs petites procédures dont chacune a une utilité bien précise.

La procédure GENMOD de SAS est une procédure visant uniquement à traiter des modèles linéaires généralisés alors que Glim4 (qui signifie **G**eneralised **L**inear **I**nteractive **M**odeling) est un logiciel à part entière.

Nous allons donc nous pencher sur les modèles linéaires généralisés et comparer ces deux logiciels sur ce point.

Pour cette étude, nous allons dans un premier temps rappeler rapidement ce que sont les modèles linéaires généralisés, ensuite comparer les deux logiciels et enfin donner des exemples.

Chapitre 1

Les Modèles Linéaires Généralisés

Ce chapitre n'est qu'un bref rappel de cours pour montrer les notations employées.

1.1 Définitions

- Un **modèle linéaire généralisé** est la donnée des observations y_1, \dots, y_n de n variables aléatoires réelles Y_1, \dots, Y_n indépendantes et de densités respectives :

$$f_i(y, \beta, \zeta) = \mathbb{1}_U e^{\frac{1}{\zeta} [y\theta_i - b(\theta_i)] + c(\zeta, y)} \quad , \quad i = 1, \dots, n$$

Avec:

$$\theta_i = a(\langle t_i, \beta \rangle)$$

$\beta \in \mathbb{R}^p$ et $\zeta \in \mathbb{R}^+$ inconnus

t_i, a, b, c applications connues

a injective

a et b possédant des dérivées secondes

U indépendant de β et de ζ .

- $\langle t_i, \beta \rangle$ est le **prédicteur linéaire** pour l'observation y_i .
- $a^{-1} \circ b'^{-1}$ est la **fonction de lien l**.
- Si $a = Id$ alors $l = b'^{-1}$ est la **fonction de lien canonique**.
- $b'' \circ b'^{-1}$ est la **fonction variance**.

- ζ est le **paramètre d'échelle**.
- $\hat{r}_i = \frac{y_i - \hat{y}_i}{\sqrt{b''(\hat{\theta}_i)}}$ est le **résidu de Pearson** (avec \hat{y}_i l'estimation de y_i et $\hat{\theta}_i$ celle de θ_i).
- $\chi^2 P = \sum_i (\hat{r}_i)^2$ est le χ^2 **de Pearson**.
- Etant donné un Modèle Linéaire Généralisé défini comme précédemment, on lui associe le **modèle saturé** défini par:
 Y_1, \dots, Y_n sont indépendants.
 Y_i a pour densité:
 $f_i(y, \theta, \zeta) = \mathbb{I}_U e^{\frac{1}{\zeta}(y\theta_i - b(\theta_i)) + c(\zeta, y)}$, $i = 1, \dots, n$.
Le paramètre inconnu est donc $(\theta_1, \dots, \theta_n, \zeta)$.
La **Log-vraisemblance** du modèle saturé est donc:
 $L(y, \theta, \zeta) = \frac{1}{\zeta} \sum_i [y_i \theta_i - b(\theta_i)] + \sum_i c(\zeta, y_i)$.
L'estimation de θ par le maximum de vraisemblance est $\hat{\theta}_i = b'^{-1}(y_i)$.
De plus, on pose $\hat{L} = L(y, \hat{\theta}, \zeta) = \frac{1}{\zeta} \sum_i (y_i \hat{\theta}_i - b(\hat{\theta}_i))$.
Etant donné donc un Modèle Linéaire Généralisé avec ζ inconnu, on appelle **déviance** du modèle:
Dev = $2\zeta(\hat{L} - L(y, \hat{\beta}, \zeta))$.
- Si ζ est connu, la **déviance normée (scaled déviance)** est:
S.Dev = $2(\hat{L} - L(y, \hat{\beta}))$.

1.2 Propriétés

- $\mu_i = E(Y_i) = b'(\theta_i)$ est la **moyenne**.
- $\text{Var}(Y_i) = \zeta b''(\theta_i)$ est la **variance**.
- L'**intervalle de confiance** de β_j , de sécurité approximative $(1 - \alpha)$, est $[\hat{\beta}_j - l * s.e.; \hat{\beta}_j + l * s.e.]$ où l est le $(1 - \frac{\alpha}{2})$ quantile de $N(0,1)$ et $s.e$ est l'erreur standard de $\hat{\beta}_j$.
- **Les Tests:**
Soit $H_0 = \{\beta/Q\beta = 0\} * \mathbb{R}^+$ avec Q matrice $q * p$ de rang q .

-Le **Test du Rapport de Vraisemblance** est:

$$\begin{aligned} T_1 &= 2[L(y, \hat{\beta}, \hat{\zeta}) - \sup_{H_0} L(y, \beta, \zeta)] \\ &= 2[L(y, \hat{\beta}, \hat{\zeta}) - L(y, \hat{\beta}_0, \hat{\zeta}_0)]. \end{aligned}$$

-Le **Test de Wald** est:

$$T_2 = (Q^T \hat{\beta})^T (Q^T \hat{\Sigma} Q)^{-1} (Q^T \hat{\beta}) \text{ où } \hat{\Sigma} \text{ est l'estimation de la matrice de variance-covariance de } \beta.$$

REMARQUE: Le **Test de Wald** pour $H_0 = \{\beta/\beta_j = 0\} * \mathbb{R}^+$ est $T_2 = [\frac{\hat{\beta}_j}{s.e.}]^2$.

Quel que soit le test T ci dessus, H_0 est rejeté $\Leftrightarrow T > (1 - \alpha)$ quantile de la distribution χ_q^2 .

1.3 Exemples

1.3.1 Modèle Gaussien

- $Y_i \sim N(g(< t_i, \beta >), \sigma^2) \Rightarrow f_i(y, \beta, \zeta) = e^{-\frac{1}{2\sigma^2}(y-g(< t_i, \beta >))^2}$.
- $\theta_i = g(< t_i, \beta >) \Rightarrow a = g, b(\theta_i) = \frac{1}{2}\theta_i^2$ et $\zeta = \sigma^2$.
- $\mu_i = E(Y_i) = g(< t_i, \beta >)$ et $Var(Y_i) = \sigma^2 = \zeta$.
- fonction variance : 1 et fonction de lien : g^{-1} .
- $\hat{r}_i = y_i - \hat{y}_i$ avec $\hat{y}_i = \hat{\mu}_i = g(< t_i, \hat{\beta} >)$ et $Dev = \sum_i (y_i - \hat{y}_i)^2$.

1.3.2 Modèle Gamma

- $Y_i \sim G(\lambda_i, \gamma) \Rightarrow f_i(y_i) = \mathbb{1}_{\mathbb{R}^+} \frac{\lambda_i^\gamma}{\Gamma(\gamma)} y_i^{\gamma-1} e^{-\lambda_i y}$ avec $\lambda_i > 0$ et $\gamma > 0$.
- $\theta_i = -\frac{\lambda_i}{\gamma}, b(\theta_i) = -\log(-\theta_i)$ et $\zeta = \frac{1}{\gamma}$.
- $E(Y_i) = \frac{\gamma}{\lambda_i}$ et $Var(Y_i) = \frac{\gamma}{\lambda_i^2}$.
- Fonction variance: $\mu \rightarrow \mu^2$.

- Liens: –Modèle linéaire: $\mu_i = \langle t_i, \beta \rangle$ $l = \text{Id}$.
–Modèle Log-linéaire: $\mu_i = e^{\langle t_i, \beta \rangle}$ $l = \text{Log}$.
–Modèle inverse: $\mu_i = -\frac{1}{\langle t_i, \beta \rangle}$ lien canonique = –Inv
(où Inv est la fonction inverse).

- $\hat{r}_i = \frac{y_i - \hat{y}_i}{y_i}$ où $\hat{y}_i = \frac{\hat{\lambda}_i}{\lambda_i} = \hat{\mu}_i$ et $Dev = 2 \sum_i ((\frac{y_i}{\hat{y}_i} - 1) + \log(\frac{y_i}{\hat{y}_i}))$.

1.3.3 Modèle Poissonien

- $Y_i \sim \mathcal{P}(\lambda_i) \Rightarrow f_i(y, \beta) = e^{-\lambda_i} \frac{\lambda_i^y}{y!}$.
- $\theta_i = \log(\lambda_i) \Rightarrow b(\theta_i) = e^{\theta_i}$ et $\zeta = 1$.
- $Var(Y_i) = \mu_i = \lambda_i$.
- fonction variance: Id.
- Liens: –Modèle linéaire: $\lambda_i = \langle t_i, \beta \rangle$ $l = \text{Id}$.
–Modèle Log-linéaire: $\lambda_i = e^{\langle t_i, \beta \rangle}$ lien canonique = Log.
- $\hat{r}_i = \frac{y_i - \hat{y}_i}{\sqrt{\hat{y}_i}}$ et $\hat{y}_i = \hat{\lambda}_i$ et $S.Dev = 2 \sum_i (y_i \log \frac{y_i}{\hat{y}_i} + \hat{y}_i - y_i)$.

1.3.4 Modèle Binomial

- $Y_i \sim B(n_i, p_i) \Rightarrow f_i(y, \beta) = \frac{n_i!}{y!(n_i-y)!} p_i^y (1-p_i)^{n_i-y}$ avec $y \in \{0, \dots, n_i\}$.
- $\theta_i = \log \frac{p_i}{1-p_i}$, $b(\theta_i) = n_i \log(1 + e^{\theta_i})$ et $\zeta = 1$.
- $Var(Y_i) = n_i p_i (1-p_i)$ et $\mu_i = n_i p_i$.
- Liens:
–Modèle Logit: $p_i = \frac{e^{\langle t_i, \beta \rangle}}{1 + e^{\langle t_i, \beta \rangle}}$; lien canonique : $l(\mu_i) = \log \frac{\mu_i}{n_i - \mu_i}$.
–Modèle probit: $p_i = \Phi(\langle t_i, \beta \rangle)$ $l(\mu_i) = \Phi^{-1}(\frac{\mu_i}{n_i})$.
–Modèle Complémentaire Log-Log:
 $p_i = e^{-e^{\langle t_i, \beta \rangle}}$ et $l(\mu_i) = \log(-\log \frac{\mu_i}{n_i})$.
- $\hat{r}_i = \frac{y_i - n_i \hat{p}_i}{\sqrt{n_i \hat{p}_i (1 - \hat{p}_i)}}$, $\hat{y}_i = n_i \hat{p}_i$ et $S.Dev = 2 \sum_i (y_i \log \frac{y_i}{\hat{y}_i} + (n_i - y_i) \log \frac{n_i - y_i}{n_i - \hat{y}_i})$.

Chapitre 2

Comparaison de Glim4 et de SAS-GENMOD

2.1 Brève explication de Glim4 et SAS-GENMOD

Brève explication sur le fonctionnement des logiciels

- Glim4 (tout comme SAS) ne différencie pas les minuscules des majuscules. Nous pouvons employer des abréviations des directives de Glim4. Les abréviations minimales sont indiquées en majuscule dans ce manuel.
- Dans Glim4, un '\$' en entrée représente une directive (une opération), un '\$' en sortie signifie que l'on en a fini avec la directive dont on vient de se servir.
Un '%' représente un système (une fonction, etc...)
Un ':' représente une répétition.
- Dans SAS, un ';' représente la fin d'une opération.
Un '/' représente les options que l'opération va exécuter.

Lancement

- Pour lancer SAS, sous UNIX, il y a deux méthodes:
 - Taper 'sas' et travailler sur les fenêtres qui s'ouvrent:
 - "SAS: PROGRAM EDITOR" sert à taper le programme,

- "SAS: OUTPUT" représente les sorties,
- "SAS: LOG" décrit l'exécution du programme.
- Si l'on a déjà écrit et sauvegardé le programme à exécuter, taper '**sas**' suivi du nom du fichier qui contient le programme.
Deux fichiers vont se créer:

- Un fichier du même nom que celui du programme avec l'extension **.lst** va se créer, ce sera le fichier des sorties,
- Un fichier du même nom que celui du programme avec l'extension **.log** va aussi se créer, ce sera le fichier de compilation (des erreurs, etc...).

- Pour lancer Glim4, il faut taper '**glim**' et contrairement à SAS, il n'y a aucune fenêtre qui s'ouvre.

Sortie

- Pour sortir de SAS, il suffit de cliquer sur "**File - Exit - OK**".
- Pour sortir de Glim4, il faut taper '**\$\$Top**'.

Sauvegarde de fichiers

- Dans SAS, pour sauvegarder un fichier, il suffit de cliquer sur "**File - Save as - Write to file**".
- Dans Glim4, une sauvegarde automatique des programmes se fait.
Le fichier des entrées-sorties est le fichier **Glim4.log**, celui des entrées seulement est le fichier **Glim4.jou**.
On peut de plus sauvegarder un programme.
Pour cela, il faut taper '**\$Output 'nom du fichier'\$**' avant de taper ce programme.

Récupération des fichiers

- Dans SAS, pour récupérer un fichier, il suffit de cliquer sur ”**File - Open - Read file**”.
- Dans Glim4, pour récupérer un fichier, il faut taper ’**\$Input** *nom du fichier*’.

Sauvegarde de données

- Dans SAS, pour sauvegarder les données obtenues dans une table (pour pouvoir s’en servir dans une autre procédure), on utilise l’option ’**MAKE**’ (voir l’exemple sur les graphiques).

Ce que cette option permet d’obtenir est représenté dans le tableau ’tab. 2.1’, page suivante.

ATTENTION: Les tables que sauvegarde ’**MAKE**’ doivent être dans la description du modèle.

Par exemple, si l’on précise ’**OBSTATS**’, on obtient en sortie la table ’**Observation Statistics**’. Pour sauvegarder la table ’**OBSTATS**’, il faut signaler ’**OBSTATS**’ dans la déclaration du modèle et écrire ”**MAKE** ’**OBSTATS**’ **out=op**”. Ainsi, la table op sera créée et contiendra la table dont le nom en sortie est ’**Observation Statistics**’.

- Dans Glim4, des sauvegardes sont prévues pour les données, certaines doivent être extraites à l’aide de ’**\$EXtract** *nom*’, d’autres sont déjà disponibles et commencent par un ’%’.

On peut de plus sauvegarder les données que l’on veut en tapant ’**\$DUmp** *nom du fichier*’, et les récupérer en tapant ’**\$REStore** *nom du fichier*’.

Affichage à l’écran

- Dans SAS, on peut afficher à l’aide de la procédure ’**PRINT**’. Cette procédure permet d’afficher tout ce qui a été sauvegardé dans une table (voir applications).

Nom de la table	Nom des données en sortie
ClassLevels	class level information
Coefficients#	coefficients for contrast ' <i>label</i> '
Contrast	likelihood ratio statistics for contrasts
Corr	estimated correlation matrix
Cov	estimated covariance matrix
IterCont	iteration history for contrasts
IterLRCI	iteration history for LR confidence intervals
IterParms	iteration history for parameter estimates
IterType3	iteration history for Type 3 contrasts
Lagrange	Lagrange multiplier statistics
LastGrad	last evaluation of the Gradient
LastHess	last evaluation of the Hessian
LRCI	LR confidence intervals for parameters
ModFit	criteria for assessing model fit
ModInfo	model information
ObStats	observation statistics
ParmEst	analysis of parameter estimates
ParmInfo	parameter information
Type1	LR statistics for Type 1 analysis
Type3LR	LR statistics for Type 3 analysis
Type3W	Wald statistics for Type 3 analysis
WaldCI	normal confidence intervals for parameters

TAB. 2.1 - Possibilités pour l'option '**MAKE**' de SAS-GENMOD

- Dans Glim4, on peut afficher à l'aide de '**\$P**rint *variable*\$' et '**\$L**ook *variable*\$'.

Aide

- Dans SAS, on peut avoir une aide en cliquant sur "**Help**".
- Dans Glim4, on peut avoir une aide de ce que l'on veut en tapant '**\$M**ANual *Directive ou %...\$*'. (taper '**\$man info**\$' pour avoir la liste des directives). On peut aussi taper '**\$H**elp\$' juste après avoir fait une erreur pour avoir une explication détaillée. Pour avoir tous les renseignements sur ce logiciel (liste des directives, etc...), taper '**\$man**\$'.

Commentaires

- Dans SAS-GENMOD, les commentaires se font avec `'/*...*/'`.
- Dans Glim4, ils se font grâce à la directive `'$Comment commentaires$'` ou `'!'`.

Graphiques

- Dans SAS, on peut faire des graphiques à l'aide de la procédure `'PLOT'` (voir les applications).
- Dans Glim4, pour les faire, il suffit de taper `'$Plot ...$'`.

Rappeler le programme précédent, effacer le programme

- Dans SAS, pour rappeler le programme précédent, il suffit de cliquer sur `"Local - Recall text"`.
Pour effacer le programme, il faut cliquer sur `"Edit - Clear text"`.
- Dans Glim4, le fonctionnement est différent. Il est donc inutile de préciser ces possibilités.

2.2 Comparaison pour les Modèles Linéaires Généralisés

2.2.1 Avant-Propos

Entrée des données

- Dans Glim4, il faut définir le nombre de données par `'$UNits entier$'`. Ensuite, il faut déclarer le nom des variables par `'$Data variables$'`.

Puis, il faut entrer les données par **'\$Read nombres\$'**.

Enfin, il faut déclarer la réponse par **'\$Yvariate variable\$'**.

Il y a donc 3 possibilités:

- Déclarer les données dans le programme de la manière décrite ci-dessus.
- Si les données seules (sans directives) se trouvent dans un fichier à part, déclarer **'\$UNits', '\$Data', '\$Input 'nom du fichier\$'** (qui joue le rôle de **'\$Read'** et de **'\$Input'**) et enfin **'\$Yvariate'**.
- Si un fichier à part contient les données ainsi que les directives **'\$UNits', '\$Data', '\$Read'** et **'\$RETurn\$'** pour fermer le fichier (toute autre directive est facultative), alors, il faut faire **'\$Input 'nom du fichier\$'** pour importer celui ci.

- Dans SAS-GENMOD, on doit écrire **'DATA nom;'** puis **'INPUT noms des variables;'** pour déclarer le nom des variables, **'CARDS;'** (qui joue le même rôle que **'\$Read'** dans Glim4) pour entrer les données (suivies d'un **','**).

Si les données seules sont dans un fichier, taper **'INFILE "nom du fichier";'** entre **'DATA'** et **'INPUT'**.

'INFILE' joue le rôle de **'CARDS'** et d'insertion de fichier.

La déclaration de la réponse se fera dans la déclaration du modèle (**'MODEL y=.../...;'**).

Si plusieurs observations sont sur la même ligne, mettre **'@@'** à la fin de la commande **'INPUT'** (avant le **','**).

Spécification des effets

- Dans SAS, on peut spécifier les effets soit en les décrivant tous, soit en spécifiant jusqu'à quel ordre on les veut.

Par exemple, si a, b et c sont des facteurs, plutôt que de taper **'1 + a + b + c + a * b + a * c + b * c + a * b * c'**, il suffit de taper **'a|b|c'**; plutôt que d'écrire **'1 + a + b + c + a * b + a * c + b * c'**, il suffit d'écrire **'a|b|c@2'**.

- Dans Glim4, on peut faire la même chose que dans SAS-GENMOD mais au lieu de taper **'**'**, il faut taper **'.'** et au lieu d'écrire **'|'**, il faut écrire **'*'**. Le **'@'** de SAS n'est hélas pas représenté mais il y a d'autres compensations comme nous le verrons plus loin.

Par conséquent, au lieu de taper **'1 + a + b + c + a.b + a.c + b.c + a.b.c'**, il suffit de taper **'a * b * c'**, de même, au lieu d'écrire **'1 + a + b + c + a.b + a.c + b.c'**,

il suffit d'écrire ' $a * b * c - a.b.c$ '.

De plus, les autres opérations possibles dans Glim4 sont:

$$a/b = a + a.b.$$

$$(a + b) **2 = (a + b) * (a + b).$$

$$a * b * c - *a = b + c + b.c.$$

$$a * b * c - /a = a + b + c + b.c.$$

Paramétrisation

- SAS paramétrise par rapport au dernier niveau.
- Glim4 paramétrise par rapport au premier niveau.

Exécution des programmes

- Dans SAS, pour faire exécuter un programme, il suffit de cliquer sur "**Local - Submit**".

L'exécution seule d'un programme de SAS donne beaucoup de choses contrairement à Glim4:

- On peut voir les données du modèle exécuté dans la table '**Model Information**' obtenue par simple déclaration du modèle.
- La table '**Class Level Information**' obtenue de la même manière donne les facteurs et leurs niveaux.
- La table '**Parameter Information**' obtenue aussi par simple déclaration du modèle donne les paramètres indépendants pour SAS.

- Dans Glim4, l'interactivité permet une exécution directe à l'aide de '**\$Fit Modèle\$**'.

L'exécution seule d'un programme de Glim4 ne donne que la déviance et le degré de liberté.

Pour avoir toutes les informations du modèle, il suffit de taper '**\$Display m\$**'.

Pour obtenir les paramètres indépendants, il suffit de taper '**\$Display p\$**'.

Autres remarques

Facteurs:

- Il faut dans SAS-GENMOD tout comme dans Glim4 définir les facteurs. Cependant, il est inutile dans SAS-GENMOD de définir le niveau des facteurs.

Par contre, dans Glim4, il faut le définir (Exemple: '**CLASS f**' pour définir le facteur f dans SAS-GENMOD, '**\$Factor f 3\$**' dans Glim4 pour définir le facteur f ayant 3 niveaux).

Données non numériques:

- Si l'on veut raisonner sur des données non numériques, on peut le faire avec SAS-GENMOD.

Pour Glim4, on doit remplacer ces données par des données numériques.

Voici un exemple pour SAS:

```
data insure;
input n c car$ age;
500 42 small 1
1200 37 medium 1
100 1 large 1
400 101 small 2;
```

Modèle binomial:

- Pour le modèle binomial, il faut déclarer le nombre d'évènements et le nombre d'essais.

Dans Glim4, on précise le nombre d'essais dans la déclaration de la distribution ('**\$Error b n\$**').

Dans SAS-GENMOD, on le signale dans la déclaration du modèle avec un / ('**MODEL y/n=...**').

Ordre:

- ATTENTION: Il y a un ordre important à respecter dans SAS-GENMOD:

```
proc genmod;
class...;
autres options;
model...;
contrast;
run;
```

2.2.2 Intercept et paramètre d'échelle

Intercept

- Dans SAS-GENMOD, il existe plusieurs options pour l'intercept:
 - Si l'on tape '**NOINT**', le modèle est traité sans intercept,
 - Si l'on tape '**INTERCEPT=nombre**', la recherche de l'intercept commence à *nombre*,
 - Si l'on tape '**INTERCEPT=nombre**' ainsi que '**NOINT**', l'intercept est fixé à *nombre*.
- Dans Glim4, on ne peut pas intervenir sur l'intercept.

Paramètre d'échelle

- Dans SAS:
ATTENTION: Le paramètre d'échelle ne correspond pas toujours à ζ . Aussi, à l'avenir, nous écrirons ζ pour parler de ζ et nous parlerons de paramètre d'échelle pour le scale parameter.

Dans SAS-GENMOD, ζ est ainsi déterminé:

NOSCALE	SCALE=VALEUR	Action
Présent	Présent	ζ fixé à <i>VALEUR</i>
Présent	Absent	ζ fixé à 1
Absent	Absent	ζ déterminé par la méthode du maximum de Vraisemblance
Absent	Présent	ζ déterminé par la méthode du maximum de Vraisemblance, commençant par le point de départ: <i>VALEUR</i>

TAB. 2.2 - Estimations de ζ pour SAS

Si l'option '**DSCALE**' est spécifiée, ζ est déterminé par le rapport de la déviance sur le degré de liberté.

Si l'option '**PSCALE**' est spécifiée, ζ est déterminé par le rapport du χ^2 de Pearson sur le degré de liberté.

Une fois ζ estimé, le paramètre d'échelle de SAS (**Scale Parameter**) est ainsi déterminé:

Distribution	Paramètre d'Echelle
Gaussien	$\sqrt{\zeta}$
Gamma	$\frac{1}{\zeta}$
Poisson	$\sqrt{\zeta}$
Binomial	$\sqrt{\zeta}$
Gaussien Inverse	$\sqrt{\zeta}$

TAB. 2.3 - Paramètres d'échelle pour SAS

- Dans Glim4, le paramètre d'échelle est ζ , estimé par le quotient de la déviance sur le degré de liberté.
Si '**\$Scale Réel\$**' est spécifié, le paramètre d'échelle est fixé à ce réel.
Le paramètre d'échelle est conservé dans '**%SC**'.

2.2.3 Liens et Distributions

- Dans SAS , les abréviations sont:

Distribution	DIST D ERROR ERR=	Lien par défaut
Gaussien	NORMAL NOR N	Identité
Gamma	GAMMA GAM G	Inverse(Puissance -1)
Poisson	POISSON POI P	Log
Binomial	BINOMIAL BIN B	Logit
Gaussien Inverse	IGAUSSIAN IG	Puissance -2

TAB. 2.4 - Distributions dans SAS-GENMOD

Lien	LINK=
Logit	LOGIT
Probit	PROBIT
Identité	IDENTITY ID
Complémentaire Log-Log	CLOGLOG CLL
Log	LOG
Puissance λ	POWER(λ) POW(λ)

TAB. 2.5 - *Liens dans SAS-GENMOD*

Supposons que dans SAS-GENMOD, on veuille se créer son propre lien, on peut le faire avec les options '**FWDLINK**' et '**INVLINK**'. Ces options sont définies en fonction de '**_MEAN_**', la moyenne, '**_XBETA_**', le prédicteur linéaire, '**_RESP_**', la réponse du modèle. Une fois créé, il ne faut pas écrire le lien dans la déclaration du modèle.

Par exemple, si l'on veut se créer le lien Log, il suffit de taper:

```
proc genmod;
...
fwdlink link=log(_MEAN_);
invlink ilink=exp(_XBETA_);
model ... / err=poisson
...;
run;
```

Supposons maintenant que l'on veuille se créer sa propre distribution, les options '**VARIANCE**' et '**DEVIANCE**' sont là pour ça. On doit définir la variance et la déviance du modèle en fonction de '**_RESP_**', de '**_XBETA_**' et de '**_MEAN_**' tout comme quand on définit la nouvelle fonction lien.

Par exemple, si l'on veut se créer la distribution de Poisson, il suffit de taper:

```
proc genmod;
class...;
a=_MEAN_;
y=_RESP_;
d=2*(y*log(y/a)-y-a));
variance var=a;
deviance dev=d;
model ...;
run;
```

- Dans Glim4, les abréviations sont:

Distribution	\$Error	Lien par défaut
Gaussien	N	I
Gamma	G	R
Poisson	P	L
Binômial	B	G
Gaussien Inverse	I	Q
Macro	O (Own)	A Définir

TAB. 2.6 - *Distributions dans Glim4*

Lien	\$Link
Logit	G
Probit	P
Identité	I
Complémentaire Log-Log	C
Log	L
Inverse (Puissance -1)	R
Racine carrée	S
Inverse Quadratique (Puissance -2)	Q
Exponentielle	E
Macro	O (Own)

TAB. 2.7 - *Liens dans Glim4*

Dans Glim4, pour déclarer sa propre distribution, il suffit d'initialiser '%VA' en fonction de '%FV' et '%DI' en fonction de '%ETA' et de '%FV'. Supposons que, dans Glim4, on veuille se créer la distribution de Poisson, la procédure est la suivante:

```
$Macro POISSON
$CALculate %VA=%FV
$CALculate %DI=2*(%YV*%log(%YV/%FV)-(YV-%FV))
$Endmac
```

Ensuite, pour lancer la macro, il suffit de taper:

```
$Error Own POISSON$
```

Pour se créer son propre lien, il suffit d'initialiser '%FV' en fonction de '%ETA' et '%DR' en fonction de '%FV' ou de '%ETA'.

Il faut de plus déclarer le point de départ pour l'algorithme avec '\$INITial macro' où l'on initialise '%ETA' en fonction de '%YV'.

Supposons maintenant que l'on veuille se créer le lien Log, la procédure est la suivante:

```
$Macro LOGLINK
$CALculate %FV=%exp(%ETA)
$CALculate %DR=1/%FV
$Endmac
```

```
$Macro LOGINIT
$CALculate %ETA=%log(%YV+0.5)
$Endmac
```

De même que pour la macro de la distribution, pour lancer le lien, il suffit de taper:

```
$Error p
$Link Own LOGLINK
$INITial LOGINIT
$Fit$
```

REMARQUE: Par défaut, les deux logiciels prennent la distribution normale.

2.2.4 Paramètres Estimés et Erreurs Standards

- Dans SAS, les paramètres estimés et leurs erreurs standards sont obtenus par la déclaration du modèle.

- Dans Glim4, il faut taper '\$Display a\$' ou '\$Display e\$' ou encore '\$Display u\$'.

Les valeurs des paramètres estimés sont stockées dans '%PE' et celles des erreurs standards dans '%SE'.

ATTENTION: Les paramètres ne sont pas les mêmes dans SAS et dans Glim4!

2.2.5 Résidus, Valeurs prédites et Erreurs standards

Résidus

- Dans SAS, pour obtenir les résidus, il faut taper en entrée '**OBSTATS**' dans la déclaration du modèle.

En sortie, on les trouve dans '**Observation Statistics**' sous les noms de '**Resraw**', le "Résidu Pur" r_i , '**Reschi**' le "Résidu de Pearson" \hat{r}_i et '**Resdev**' le "Résidu déviance" r_{d_i} , avec :

$$r_i = u_i - \mu_i \text{ et } r_{d_i} = \sqrt{d_i} \text{sign}(r_i)$$

(où d_i est la déviance pour chaque observation).

- Dans Glim4, on obtient le Résidu de Pearson en tapant '**\$Display r\$**' ou '**\$Display w\$**' après avoir écrit le modèle. Ce Résidu est stocké dans '**%RS**'.

Les autres Résidus ne peuvent s'obtenir qu'en les calculant.

Par exemple, pour obtenir le "Résidu Pur", il suffit de taper:

```
$Macro ri
$Calculate %yv-%fv
$Endmac
```

Valeurs prédites

- Dans SAS, elles s'obtiennent avec l'option '**OBSTATS**' dans la déclaration du modèle. On les trouve en sortie sous le nom '**Pred**'.
- Dans Glim4, on les obtient en tapant '**\$Display r\$**' ou '**\$Display w\$**' après avoir écrit le modèle. Ces valeurs sont stockées dans '**%FV**'.

Erreurs standards

- Dans SAS, les erreurs standards des valeurs prédites s'obtiennent aussi avec l'option '**OBSTATS**' dans la déclaration du modèle. On les trouve en sortie sous le nom '**Std**'.
- Dans Glim4, les erreurs standards des valeurs prédites ne sont pas prévues, il faut donc les créer.

2.2.6 Tests et Intervalles de confiance

Tests

- Dans SAS, le test de **Wald** pour les paramètres estimés (comme définit dans le rappel de cours) se fait automatiquement par simple exécution du modèle.

Le test de type 1 est le test du rapport de vraisemblance.

Il permet de tester le modèle de manière séquentielle (d'abord uniquement avec l'intercept, ensuite, avec un seul facteur, etc...).

Ce test s'obtient en déclarant '**TYPE 1**' dans la déclaration du modèle.

Le test de type 3 est aussi le test du rapport de vraisemblance.

Celui-ci permet de tester le modèle pour chaque effet (par exemple: effet ligne, puis effet colonne, puis interactions, etc...).

Ce test s'obtient en déclarant '**TYPE 3**' dans la déclaration du modèle.

Si l'option '**WALD**' est dans la déclaration du modèle en plus de '**TYPE 3**', c'est le test de Wald qui est exécuté).

On peut de plus tester $Q^T \beta = 0$ par le test du rapport de vraisemblance, où β est le vecteur des paramètres et Q est une matrice comme défini dans le rappel de cours. Pour tester cela, **sas** a prévu la commande '**CONTRAST**'.

On peut rajouter à cette commande '**E**' pour faire afficher la matrice en sortie, et '**WALD**' pour faire le test de Wald au lieu du test du rapport de vraisemblance.

Exemple:

```
class l c;
model...;
CONTRAST 'nom_test1' l 0 1,
           c 1 1 1;
CONTRAST 'nom_test2' l*c 0 1,
           l 1/E Wald;
```

Remarque: Les matrices non complètes sont complétées par des 0.

- Dans Glim4, il n'y a pas de Tests, il faut donc se les créer soi-même.

Intervalles de Confiance

- Dans SAS, l'option '**OBSTATS**' permet d'obtenir un intervalle de confiance des valeurs prédites.

On peut de plus obtenir des intervalles de confiances des paramètres à l'aide des options '**WALDCI**' et '**LRCI**'.

Ces intervalles de confiance sont à 95% par défaut. Cependant, on peut fixer la sécurité α avec l'option '**ALPHA**=*nombre entre 0 et 1*' (par défaut $\alpha = 0.05$).

- Dans Glim4, il n'y a pas d'intervalles de confiance, il faut donc se les créer soi-même.

Par exemple:

```
$number nval
$cal nval=%ND(0.975):
    lower=%PE-nval*%SE:
    upper=%PE+nval*%SE
$loo %SE %PE lower upper$
```

2.2.7 Corrélation et Covariance

- Dans SAS, les estimations des matrices de Variance-Covariance et de Corrélation des paramètres s'obtiennent respectivement avec les options '**COVB**' et '**CORRB**' dans la description du modèle.

- Dans Glim4, l'estimation de la matrice de Corrélation des paramètres s'obtient par '**\$Display c\$**' et l'estimation de la matrice de Variance-Covariance des paramètres par '**\$Display v\$**'.

L'estimation de la matrice de Variance-Covariance est stockée dans '**%VC**' après l'avoir extraite en tapant '**\$EXTRACT %VC\$**'.

L'estimation de la matrice de Corrélation n'est pas stockée, il faut donc la créer.

ATTENTION: Les paramètres ne sont pas les mêmes dans SAS et dans Glim4!

2.2.8 Poids et offset

Poids

Tout ce qui a été traité précédemment a été traité en supposant qu'il n'y avait pas de poids (y compris dans le rappel de cours).

Si l'on veut mettre des poids, voici la démarche à suivre:

- Dans SAS-GENMOD, si l'on veut qu'une variable représente le poids, il suffit de déclarer '**SCWGT** *variable*'.

- Dans Glim4, pour mettre des poids, il suffit de déclarer '**\$Weight** *variable*'.

Ce vecteur-poids est stocké dans '**%PW**'.

Quand on veut changer de modèle, si on ne veut plus de poids, il suffit de déclarer '**\$Weight**'.

Offset

- Dans SAS, pour déclarer l'offset, il suffit de déclarer '**OFFSET=***variable*;' dans la déclaration du modèle.

- Dans Glim4, pour mettre un offset, il suffit de déclarer '**\$Offset** *Variable*'. L'offset est ensuite stocké dans '**%OS**'.

2.2.9 Autres

χ^2 de Pearson

- Dans SAS, le χ^2 de Pearson, le scaled χ^2 de Pearson, et leurs quotients respectifs sur le degré de liberté du modèle sont toujours obtenus simplement par déclaration du modèle.

- Dans Glim4, le χ^2 de Pearson se trouve dans le working-triangle.

C'est le dernier nombre de ce working-triangle, c'est à dire, le nombre dont le terme de la colonne et celui de la ligne sont les plus élevés).

Ce working-triangle s'obtient par '**\$Display t**'.

Le χ^2 de Pearson est aussi obtenu avec '**%X2**'.

Les autres nombres ne sont pas prévus.

Prédicteur linéaire

- Dans SAS, le prédicteur linéaire est obtenu en tapant l'option '**OBS-TATS**' et est en sortie sous le nom '**Xbeta**'.

- Dans Glim4, le prédicteur linéaire est obtenu en tapant '**%LP**'.

Si l'on veut modifier le prédicteur linéaire (pour une macro, ou autre), on ne peut pas toucher à '**%LP**' mais par contre, on peut et on doit modifier '**%ETA**'.

La variance du prédicteur linéaire pour chaque observation est obtenue dans '**%VL**' après l'avoir extraite avec '**\$EXTRACT %VL\$**'.

La Dérivée de celui ci (et donc du lien) pour chaque observation est dans '**%DR**' après l'avoir extraite avec '**\$EXTRACT%DR\$**'.

Déviante

- Dans SAS, la déviante et la déviante normée du modèle sont obtenues simplement par déclaration du modèle.

Dans le test de Type I, on obtient les déviants des modèles auxquels ont été supprimés les différents facteurs et interactions (voir tests).

Pour obtenir cela, il suffit de taper '**TYPE 1**' dans la déclaration du modèle.

- Dans Glim4 aussi, la déviante du modèle (ou déviante normée si le paramètre d'échelle est 1) est toujours obtenue par simple déclaration du modèle et exécution.

De plus, dans Glim4, la déviante du modèle est stockée dans '**%DV**'.

La déviante pour chaque observation est stockée dans '**%DI**' après l'avoir extraite avec '**\$EXTRACT %DI\$**'.

Fonction variance

- Dans SAS, les valeurs de la fonction variance pour chaque observation ne sont pas prévues.

- Dans Glim4, elles sont stockées dans '**%VA**' après les avoir extraites avec '**\$EXTRACT \$VA\$**'.

Matrice du modèle

- Dans SAS, on ne peut pas consulter la matrice du modèle.
- Dans Glim4, cette matrice est stockée dans '**%DM**' après l'avoir extraite avec '**\$EXTRACT %DM\$**'.

Log-vraisemblance

- La **Log-vraisemblance** du modèle est calculée dans SAS-GENMOD par simple déclaration de ce modèle et se trouve sous le nom de '**Log-Likelihood**' mais n'est pas calculée dans Glim4.

Fréquence

- La fréquence de chaque observation peut être précisée dans SAS-GENMOD à l'aide de la commande '**FREQ**' suivie de la variable représentant la colonne des fréquences.
- Dans Glim4, cette option n'est pas prévue.

Analyse par niveau

- Dans SAS-GENMOD, l'analyse par niveau est possible à l'aide de l'option '**BY**' (par exemple, si le facteur l a deux niveaux, '**BY l**' donnera une analyse pour l=1 puis une analyse pour l=2).
- Dans Glim4, cette option n'est pas prévue, excepté avec l'option '**\$Tabulate ...**'. Cette dernière option ne permet pas une analyse complète comme dans sas, cependant l'analyse par niveau n'est qu'une option parmi d'autres de '**\$Tabulate**'.

Réponse du modèle

- Dans SAS-GENMOD, la Réponse du modèle est obtenue dans **'Value'** après avoir lancé l'option **'OBSTATS'**. Cette réponse est aussi représentée par **'_PRED_'**.
- Dans Glim4, la réponse est dans **'%YV'**.

Méthode algorithmique

- Dans SAS-GENMOD, on ne peut choisir la méthode avec laquelle le programme sera exécuté.
- Dans Glim4, on peut utiliser la méthode Gauss-Jordan ou Givens à l'aide de **'\$Method J ouG'**. De plus, on peut se créer sa propre méthode par une macro. Enfin, pour vérifier la méthode employée, il suffit de taper **'\$Display m'**.

Macros et Calculs

- Dans SAS-GENMOD, on peut faire des macros avec **'%macro nom'** pour ouvrir, **'%mend nom'** pour fermer et **'%nom'** pour appeler (voir l'aide). Les calculs se font de manière habituelle (par exemple, pour le logarithme, taper Log...etc).
- Dans Glim4, on peut aussi se créer des macro:
 - **'\$Macro nom de la macro'** pour les ouvrir,
 - **'\$Endmac\$'** pour les fermer,
 - **'\$Use nom de la macro\$'** pour les utiliser.Certaines macros sont déjà créées dans la librairie des macros. Cette librairie se trouve au CICT à **"/usr/local/Glim4"** sous le nom **'macrolib'** (taper **'\$ MANUAL LIBRARY \$'** pour avoir plus de renseignements). Les calculs se font avec la commande **'\$CALculate calcul\$'**. Les fonctions sont précédées d'un % (exemples: %cos, %sin, %abs, %exp...).

Lancer une autre étude

- Dans SAS-GENMOD, la question ne se pose pas.
- Dans Glim4, on peut éviter de sortir pour traiter de nouvelles données en tapant '**\$NEWjob\$**'.

Chapitre 3

Applications

3.1 Modèle Gaussien

Avec Glim4:

```
glim
$pag on !pour avoir la pagination
$echo on !pour avoir le detail des fichiers importes
$uni 5$
$dat x1 x2 y$
$read
1 2 33.11
2 1 2.5
3 0 0.4
4 1 0.1
5 2 0.2
$yva y$
$fit x1+x2+x1*x2$
    deviance = 59.699
residual df = 1
$loo %x2$
59.70
$fit -x1.x2$
    deviance = 138.72 (change = +79.02)
residual df = 2 (change = +1 )
$loo %x2$
138.7
```

```

$dis a$
      estimate      s.e.      parameter
1      16.74      10.58      1
2      -6.822     2.634     X1
3       9.160     4.977     X2
scale parameter 69.36

```

```

$dis r$
      unit  observed  fitted  residual
1         1   33.1100  28.2340   4.876
2         2    2.5000  12.2520  -9.752
3         3    0.4000  -3.7300   4.130
4         4    0.1000  -1.3920   1.492
5         5    0.2000   0.9460  -0.746

```

```

$dis c$
correlations between parameter estimates
1  1.0000
2 -0.7467  1.0000
3 -0.5644  0.0000  1.0000
      1      2      3

```

```

$dis v$
(co)variance matrix of parameter estimates
1      112.0
2      -20.81      6.936
3      -29.72      0.000      24.77
      1      2      3

```

```

$extract %di$

```

```

$loo %di$
      %DI
1  23.7754
2  95.1015
3  17.0569
4   2.2261
5   0.5565

```

```

$stop

```


3.2 Modèle Gamma

Avec SAS-GENMOD:

Voici ce que contient le fichier gamma11:

```
1 2 33.11
2 1 2.5
3 0 0.4
4 1 0.1
5 2 0.2
```

Programme:

```
data gamma11;
infile "gamma11";
input x1 x2 y;
proc genmod data=gamma11;
model y=x1 x2 x1*x2/ dist = gamma
      link = log
      type1;
run;
```

Sortie:

The GENMOD Procedure

Model Information

Description	Value
Data Set	WORK.GAMMA11
Distribution	GAMMA
Link Function	LOG
Dependent Variable	Y
Observations Used	5

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	1	0.5444	0.5444
Scaled Deviance	1	5.0890	5.0890
Pearson Chi-Square	1	0.4831	0.4831
Scaled Pearson X2	1	4.5157	4.5157
Log Likelihood	.	-1.1838	.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Std Err	ChiSquare	Pr>Chi
INTERCEPT	1	4.5925	1.7973	6.5289	0.0106
X1	1	-1.9416	0.5931	10.7178	0.0011
X2	1	0.0072	0.9475	0.0001	0.9940
X1*X2	1	0.3321	0.3105	1.1444	0.2847
SCALE	1	9.3475	5.8094	.	.

NOTE: The scale parameter was estimated by maximum likelihood.

LR Statistics For Type 1 Analysis

Source	Deviance	DF	ChiSquare	Pr>Chi
INTERCEPT	20.6509	0	.	.
X1	3.6002	1	10.6236	0.0011
X2	0.6669	1	8.8827	0.0029
X1*X2	0.5444	1	1.0343	0.3091

 Programme:

```
proc genmod data=gamma11;
model y=x1 x2/ dist = gamma
      link = log
      corrb
      covb
      waldci
      lrci;
run;
```

Sortie:

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	2	0.6669	0.3334
Scaled Deviance	2	5.1086	2.5543
Pearson Chi-Square	2	0.5200	0.2600
Scaled Pearson X2	2	3.9832	1.9916
Log Likelihood	.	-1.7010	.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Std Err	ChiSquare	Pr>Chi
INTERCEPT	1	2.7315	0.4252	41.2643	0.0001
X1	1	-1.3166	0.1106	141.6343	0.0001
X2	1	1.0021	0.1933	26.8660	0.0001
SCALE	1	7.6605	4.7432	.	.

NOTE: The scale parameter was estimated by maximum likelihood.

Estimated Covariance Matrix

Parameter Number	PRM1	PRM2	PRM3	Scale
PRM1	0.18081	-0.03517	-0.04099	-1.26E-10
PRM2	-0.03517	0.01224	-0.001288	3.672E-11
PRM3	-0.04099	-0.001288	0.03738	7.134E-11
Scale	-1.26E-10	3.672E-11	7.134E-11	22.49774

Estimated Correlation Matrix

Parameter Number	PRM1	PRM2	PRM3	Scale
PRM1	1.0000	-0.7477	-0.4986	-0.0000
PRM2	-0.7477	1.0000	-0.0602	0.0000
PRM3	-0.4986	-0.0602	1.0000	0.0000
Scale	-0.0000	0.0000	0.0000	1.0000

Normal Confidence Intervals For Parameters

Two-Sided Confidence Coefficient: 0.9500

Parameter	Confidence Limits	
PRM1	Lower	1.8981
PRM1	Upper	3.5649
PRM2	Lower	-1.5334
PRM2	Upper	-1.0998
PRM3	Lower	0.6232
PRM3	Upper	1.3810
Scale	Lower	-1.6359
Scale	Upper	16.9570

3.3 Modèle Poissonien

Avec Glim4:

```
glim
$pag on !pour avoir la pagination
$echo on !pour avoir le detail des fichiers importes
$uni 8$
$dat x f y$
$din pois33$
0.1 1 2
0.5 1 4
1 1 3
2 1 8
0.5 2 4
1.5 2 9
2.5 2 6
4 2 16
$fac f 2$
$yva y$
$err p$
$li i$
$fit f*x$
  scaled deviance = 3.3844 at cycle 3
  residual df = 4
$loo %x2$
  3.155
$fit -f.x$
  scaled deviance = 3.3870 (change = +0.002543) at cycle 3
  residual df = 5 (change = +1 )
$loo %x2$
  3.161
$fit -f$
  scaled deviance = 3.7234 (change = +0.3364) at cycle 3
  residual df = 6 (change = +1 )
$loo %x2$
  3.532
```

\$dis a\$

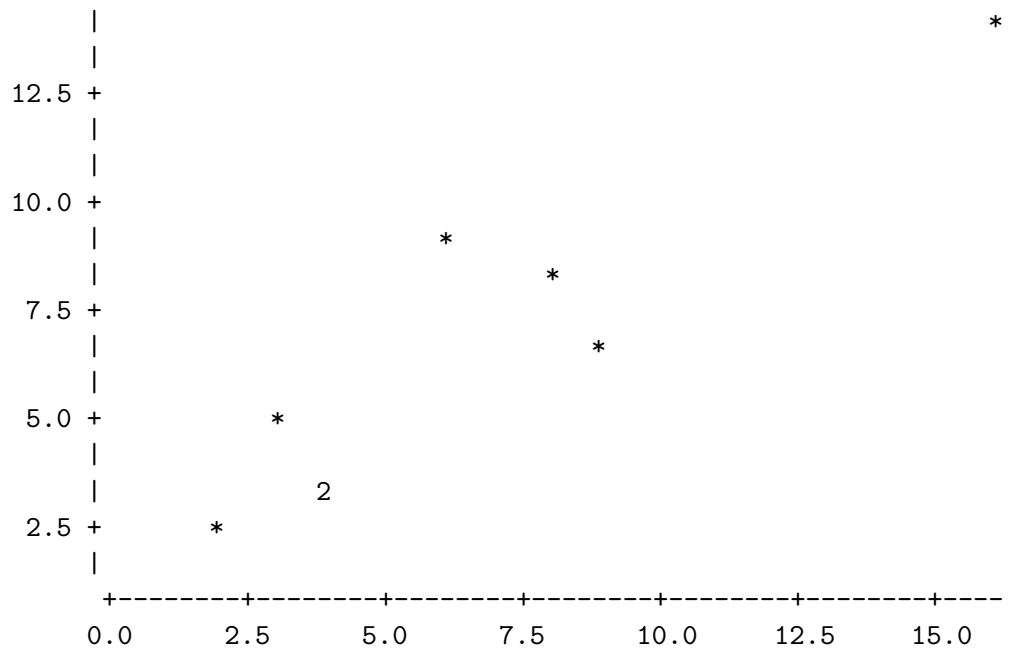
	estimate	s.e.	parameter
1	1.926	1.073	1
2	3.024	0.8049	X

scale parameter 1.000

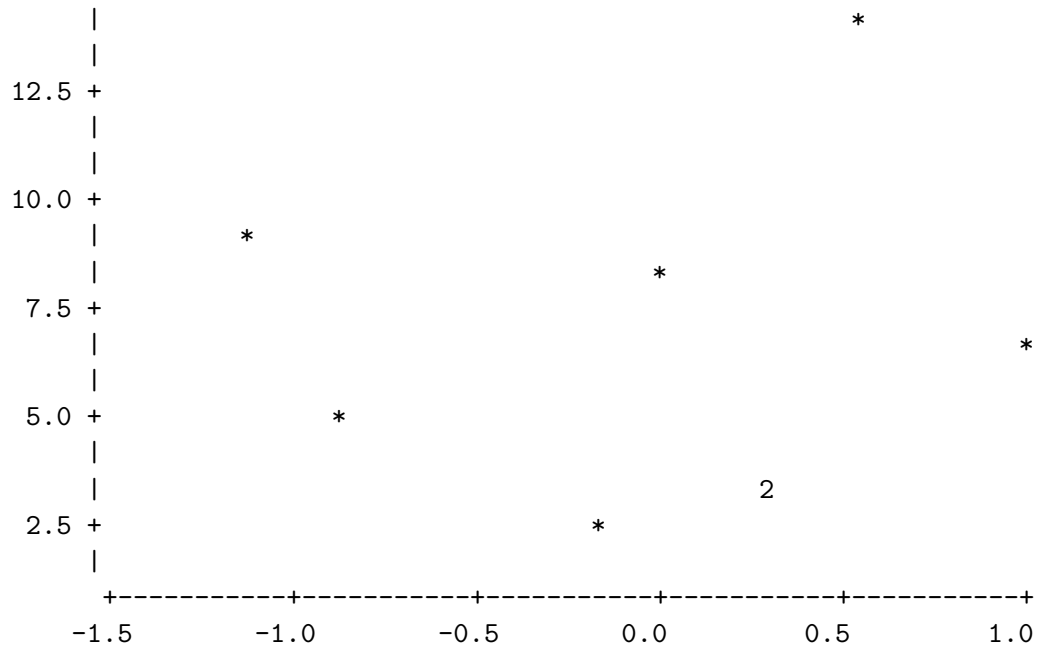
\$dis r\$

unit	observed	fitted	residual
1	2	2.228	-0.153
2	4	3.438	0.303
3	3	4.950	-0.876
4	8	7.974	0.009
5	4	3.438	0.303
6	9	6.462	0.998
7	6	9.486	-1.132
8	16	14.023	0.528

\$plo %fv %yv '*'\$



\$plo %fv %rs '*'\$



```
$extract %di$
```

```
$loo %di$
```

```
%DI
```

```
1 0.02422288
2 0.08725201
3 0.89540601
4 0.00008261
5 0.08725201
6 0.88698816
7 1.47563529
8 0.26652852
```

```
$dis c$
```

```
correlations between parameter estimates
```

```
1 1.0000
2 -0.6970 1.0000
1 2
```

```
$dis v$
```

```
(co)variance matrix of parameter estimates
```

```
1 1.151
2 -0.6019 0.6479
1 2
```

```
$stop
```

3.4 Modèle Binomial

Avec SAS-GENMOD:

Voici ce que contient le fichier bino:

```
1 1 17 35
1 2 14 20
1 3 8 25
2 1 12 30
2 2 22 40
2 3 15 50
3 1 24 30
3 2 41 60
3 3 11 20
```

Programme:

```
data bino;
infile "bino";
input l c y n;
proc genmod data=bino;
class l c;
model y/n=l c l*c/ dist = binomial
link = cll
type1;

run;
```

Sortie:

The GENMOD Procedure

Model Information

Description	Value
Data Set	WORK.BINO
Distribution	BINOMIAL
Link Function	CLL
Dependent Variable	Y
Dependent Variable	N
Observations Used	9
Number Of Events	164
Number Of Trials	310

Class Level Information

Class	Levels	Values
L	3	1 2 3
C	3	1 2 3

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	0	-0.0000	.
Scaled Deviance	0	-0.0000	.
Pearson Chi-Square	0	0.0000	.
Scaled Pearson X2	0	0.0000	.
Log Likelihood	.	-196.6287	.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Std Err	ChiSquare	Pr>Chi
INTERCEPT	1	-0.2250	0.3096	0.5283	0.4673
L	1	-0.7278	0.4716	2.3816	0.1228
L	2	-0.8059	0.4040	3.9793	0.0461
L	3	0.0000	0.0000	.	.
C	1	0.7009	0.3838	3.3347	0.0678
C	2	0.3647	0.3508	1.0809	0.2985
C	3	0.0000	0.0000	.	.
L*C	1 1	-0.1561	0.5787	0.0728	0.7874
L*C	1 2	0.7737	0.5745	1.8136	0.1781
L*C	1 3	0.0000	0.0000	.	.
L*C	2 1	-0.3417	0.5476	0.3894	0.5326
L*C	2 2	0.4412	0.4882	0.8168	0.3661
L*C	2 3	0.0000	0.0000	.	.
L*C	3 1	0.0000	0.0000	.	.
L*C	3 2	0.0000	0.0000	.	.
L*C	3 3	0.0000	0.0000	.	.
SCALE	0	1.0000	0.0000	.	.

NOTE: The scale parameter was held fixed.

LR Statistics For Type 1 Analysis

Source	Deviance	DF	ChiSquare	Pr>Chi
INTERCEPT	35.4480	0	.	.
L	15.9369	2	19.5112	0.0001
C	5.3727	2	10.5642	0.0051
L*C	-0.0000	4	5.3727	0.2512

Programme:

```
proc genmod data=bino;
class l c;
make 'obstats' out=predi;
model y/n=l c/ dist = binomial
      link = cll
      obstats
      covb
      corrb
      waldci;
run;
data new;
set predi;
proc plot data=new;
plot yvar1*reschi yvar1*pred ;
run;
```

Sortie:

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	4	5.3727	1.3432
Scaled Deviance	4	5.3727	1.3432
Pearson Chi-Square	4	5.2527	1.3132
Scaled Pearson X2	4	5.2527	1.3132
Log Likelihood	.	-199.3151	.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Std Err	ChiSquare	Pr>Chi
INTERCEPT	1	-0.3579	0.2129	2.8268	0.0927
L	1	-0.4555	0.2140	4.5322	0.0333
L	2	-0.6645	0.1946	11.6658	0.0006
L	3	0.0000	0.0000	.	.
C	1	0.5488	0.2266	5.8659	0.0154
C	2	0.6604	0.2178	9.1973	0.0024
C	3	0.0000	0.0000	.	.
SCALE	0	1.0000	0.0000	.	.

NOTE: The scale parameter was held fixed.

Estimated Covariance Matrix

Parameter Number	PRM1	PRM2	PRM3	PRM5	PRM6
PRM1	0.04531	-0.02020	-0.02171	-0.03135	-0.03681
PRM2	-0.02020	0.04578	0.01722	-0.002222	0.009396
PRM3	-0.02171	0.01722	0.03785	0.004212	0.009330
PRM5	-0.03135	-0.002222	0.004212	0.05135	0.03053
PRM6	-0.03681	0.009396	0.009330	0.03053	0.04743

Estimated Correlation Matrix

Parameter Number	PRM1	PRM2	PRM3	PRM5	PRM6
PRM1	1.0000	-0.4434	-0.5242	-0.6498	-0.7940
PRM2	-0.4434	1.0000	0.4136	-0.0458	0.2017
PRM3	-0.5242	0.4136	1.0000	0.0956	0.2202
PRM5	-0.6498	-0.0458	0.0956	1.0000	0.6187
PRM6	-0.7940	0.2017	0.2202	0.6187	1.0000

Normal Confidence Intervals For Parameters

Two-Sided Confidence Coefficient: 0.9500
 Parameter Confidence Limits

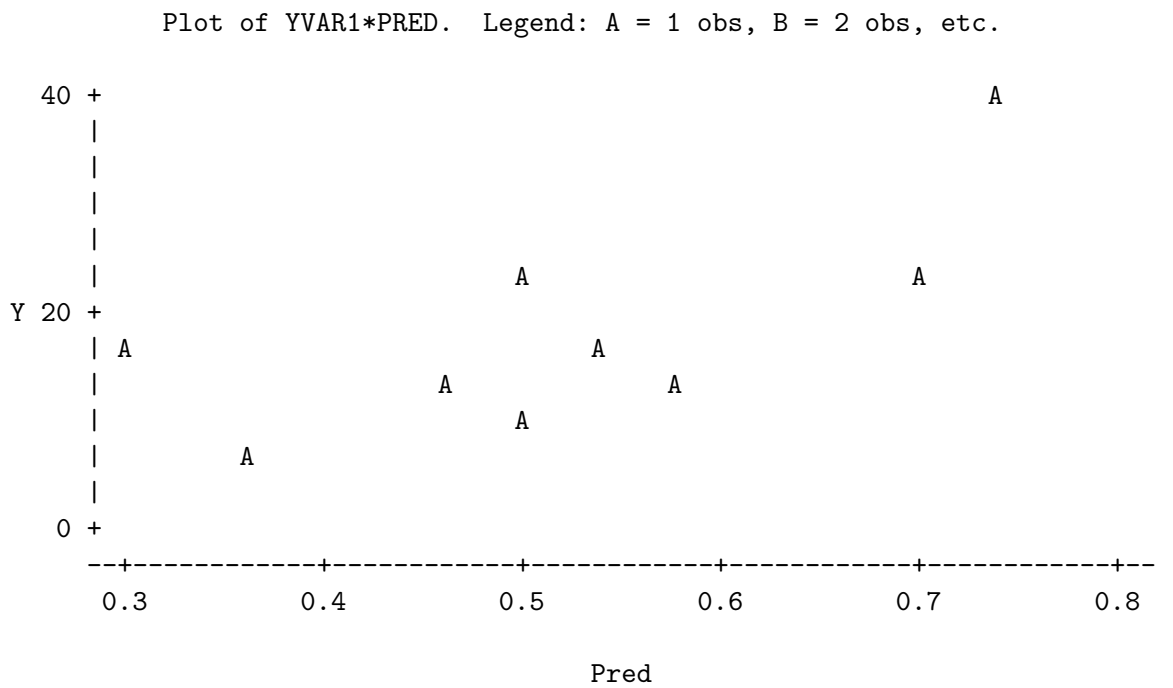
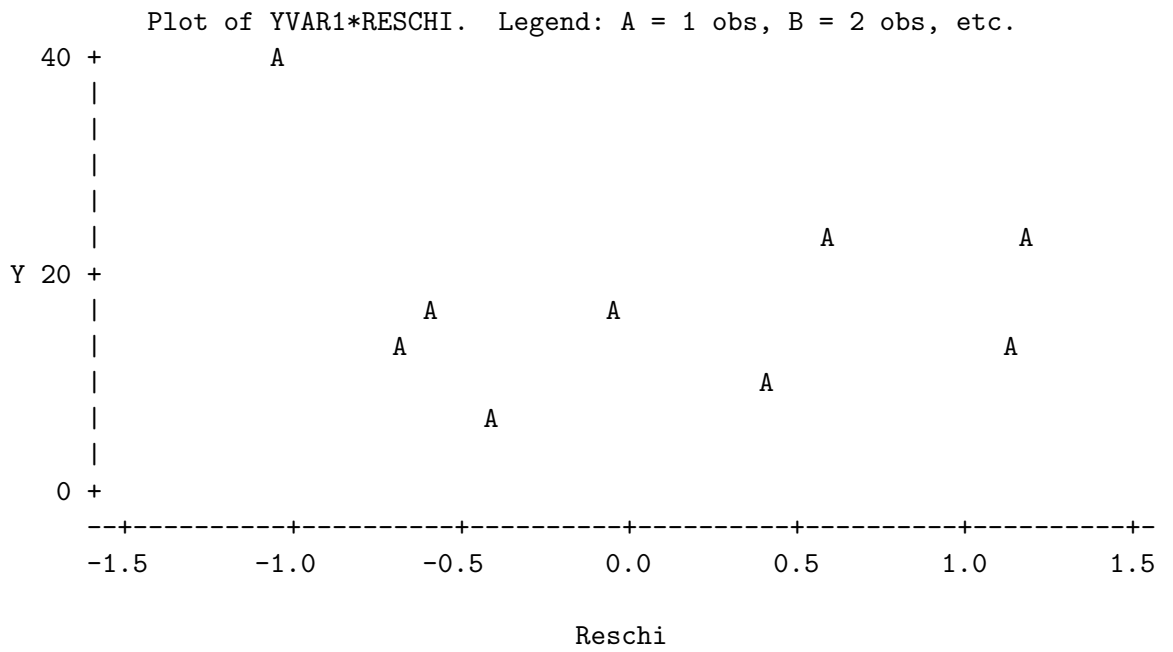
PRM1	Lower	-0.7751
PRM1	Upper	0.0593
PRM2	Lower	-0.8749
PRM2	Upper	-0.0361
PRM3	Lower	-1.0458
PRM3	Upper	-0.2832
PRM5	Lower	0.1047
PRM5	Upper	0.9930
PRM6	Lower	0.2336
PRM6	Upper	1.0873

Observation Statistics

Y	N	Pred	Xbeta	Std	HessWgt	Lower
17	35	0.5358	-0.2646	0.1869	18.7025	0.4127
14	20	0.5761	-0.1530	0.2081	9.4789	0.4349
8	25	0.3581	-0.8134	0.2252	9.0499	0.2481
12	30	0.4635	-0.4736	0.1919	14.1808	0.3479
22	40	0.5016	-0.3619	0.1795	18.4437	0.3873
15	50	0.3021	-1.0224	0.1994	14.9665	0.2160
24	30	0.7019	0.1909	0.1843	16.1985	0.5698
41	60	0.7416	0.3025	0.1383	41.6576	0.6437
11	20	0.5030	-0.3579	0.2129	9.2558	0.3691

Observation Statistics

Upper	Resraw	Reschi	Resdev
0.6694	-1.7544	-0.5946	-0.5937
0.7248	2.4788	1.1216	1.1431
0.4981	-0.9529	-0.3975	-0.4009
0.5963	-1.9064	-0.6980	-0.7011
0.6284	1.9366	0.6124	0.6130
0.4124	-0.1069	-0.0329	-0.0329
0.8240	2.9426	1.1745	1.2205
0.8305	-3.4969	-1.0313	-1.0088
0.6539	0.9402	0.4205	0.4209



Conclusion

Comme nous l'avons vu, Glim4 et SAS-GENMOD sont deux logiciels très différents.

Glim4 a pour principal avantage l'interactivité alors que SAS, pour sa part, est très complet.

On a pu constater que l'avantage de l'un est en général l'inconvénient de l'autre.

En effet, chaque logiciel a été conçu pour des raisonnements différents.

Dans Glim4, on ne raisonne que sur la déviance, alors que dans SAS-GENMOD, on peut raisonner sur la déviance mais on obtient aussi d'autres données sur lesquelles on peut raisonner.

De plus, certains résultats sont obtenus sur simple demande dans SAS-GENMOD alors que dans Glim4, il faut se les créer (par exemple les intervalles de confiance).

Pour Glim4, il est donc nécessaire de connaître un minimum de théorie alors que pour SAS-GENMOD, ce n'est pas indispensable, il suffit de savoir interpréter les résultats.

Dans SAS-GENMOD, la production des données étant importante, on a tendance à s'y perdre (Glim4 étant un logiciel interactif, on n'a pas cet inconvénient).

Dans Glim4, il n'est pas nécessaire de retaper toutes les données afin d'obtenir une nouvelle étude. Par contre, il faut le faire pour SAS-GENMOD.

En conclusion, pour ce qui est de la souplesse du logiciel, Glim4 a un énorme avantage. Par contre, en ce qui concerne l'aspect statistique, SAS-GENMOD est plus complet.

Liste des tableaux

2.1	Possibilités pour l'option 'MAKE' de SAS-GENMOD	9
2.2	Estimations de ζ pour SAS	14
2.3	Paramètres d'échelle pour SAS	15
2.4	Distributions dans SAS-GENMOD	15
2.5	Liens dans SAS-GENMOD	16
2.6	Distributions dans Glim4	17
2.7	Liens dans Glim4	17