

Régression logistiqu e : production des résultats via Statistica

Exposition à la fume de cigarette (Essenberg, 1952)

1. Feuille de données et appel au module de régression logistiqu e

The screenshot shows two windows from the Statistica software. The left window, titled 'Données : chap3_essenber...', contains a data table with the following content:

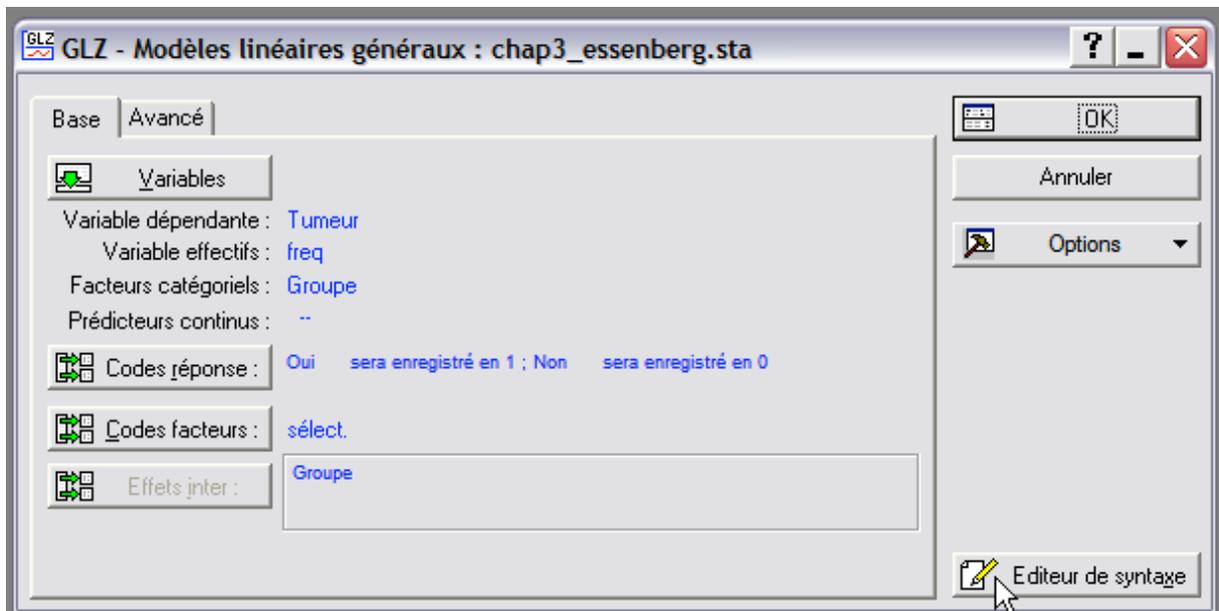
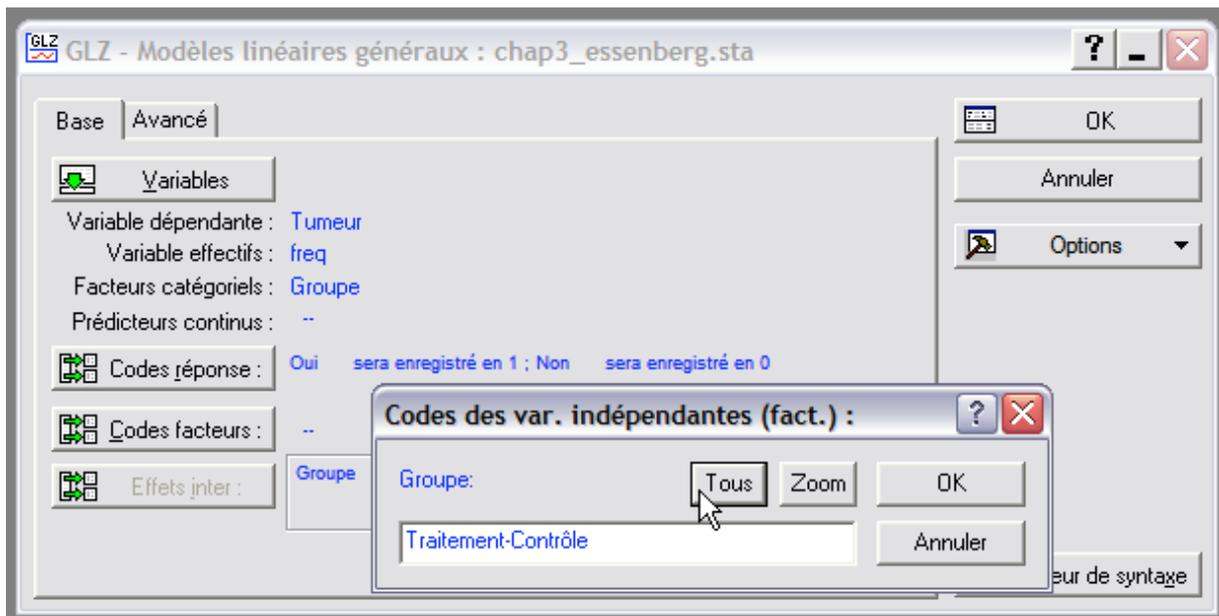
	1	2	3
	Groupe	Tumeur	freq
1	Traitemen	Oui	21
2	Contrôle	Oui	19
3	Traitemen	Non	2
4	Contrôle	Non	13

The right window, titled 'Modèles Linéaires/Non Linéaires Généralisés : chap3_essenber...', shows the 'Modèles Linéaires/Non Linéaires Généralisés' dialog box. The 'Type d'analyse' list has 'Modèles généraux spécifiques' selected. The 'Méthode de spécification' list has 'Spécifications rapides' selected. The 'Distribution' list has 'Binomiale' selected. The 'Fonctions de liaison' list has 'Logit' selected. The 'DL' dropdown is set to 'P-1'. The 'Degré' field is set to '1'.

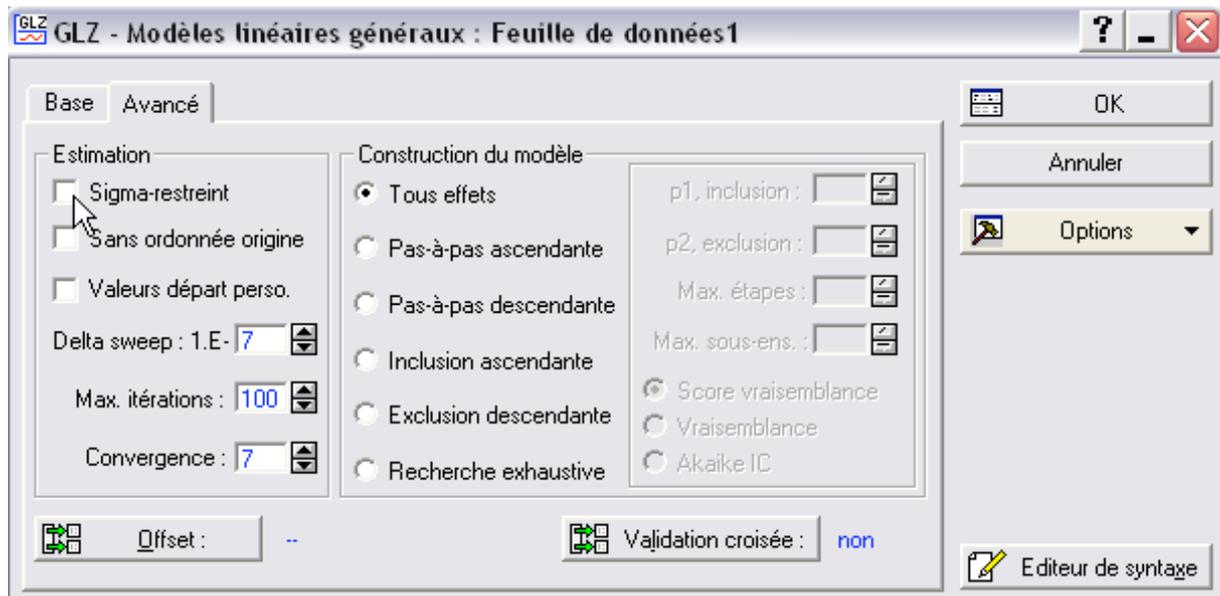
2. Spécification du modèle

The screenshot shows the 'Sélectionnez la variable dépendante et le prédicteur continu (régresseur)' dialog box. The dialog box has four columns for selecting variables. The first column is for the dependent variable, and the other three are for continuous predictors. The variables listed are '1 - Groupe', '2 - Tumeur', and '3 - freq'. The 'Variable dépendante' field contains '2'. The 'Facteurs catég.' field contains '1'. The 'Prédicteurs continus' field is empty. The 'Variable d'effectifs' field contains '3'. There are 'Résumé' and 'Zoom' buttons for each column. A checkbox at the bottom is labeled 'N'afficher que les variables appropriées'. A note on the right side of the dialog box reads: 'L'option "N'afficher que les variables appropriées" permet un filtrage préalable des listes de var. en n'affichant que les var. catégorielles et continues. Appuyez sur F1 pour plus d'informations.'

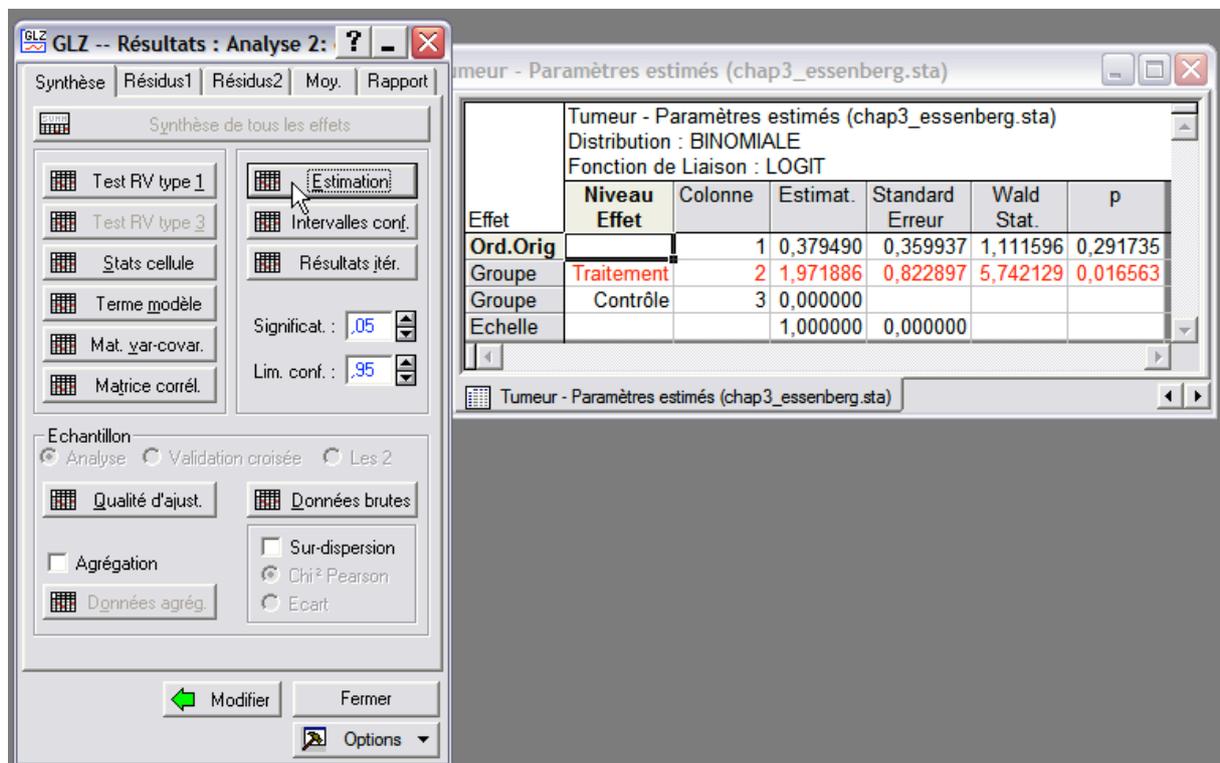
3. Codes pour les variables catégorielles réponse et explicative



4. Via l'onglet 'Avancé', décocher 'Sigma-Restreint' :



5. On peut alors examiner les estimations et erreurs standards des paramètres :



6. ... et ensuite évaluer la plausibilité que 0 est une valeur plausible pour alpha2 via la statistique G2 :

The screenshot displays the GLZ software interface. The left window, titled 'GLZ -- Résultats : Analyse 1', shows a 'Synthèse de tous les effets' section with several buttons. The 'Test RV type 1' button is highlighted with a mouse cursor. Below this, there are buttons for 'Estimation', 'Intervalles conf.', and 'Résultats itér.'. The 'Significat.' field is set to .05 and 'Lim. conf.' is set to .95. The 'Echantillon' section includes options for 'Analyse', 'Validation croisée', and 'Les 2', along with buttons for 'Qualité d'ajust.', 'Données brutes', 'Agréation', and 'Données agrég.'. The right window, titled 'Classeur1* - Tumeur - Test Vraisemblance Typ...', displays a table of results for a binomial test. The table has the following data:

Effet	Degré de Liberté	Log-Vraisbnc	Chi²	p
Ord.Orig.	1	-32,2274		
Groupe	1	-28,4100	7,634852	0,005725

On déduit de la P-valeur ($0.0057 < 5\%$) associée à G2 ($=7.63$) que 0 n'est pas une valeur plausible pour alpha2.

Influence de habitudes tabagiques des parents sur celle des enfants

En suivant la même procédure que dans l'exemple précédent, on obtient :

The screenshot shows two windows from SPSS. The left window displays the data table with columns for '1 Parents', '2 Enfant fumeur', and '3 freq'. The right window shows the 'Paramètres estimés' (Parameter Estimates) for the model.

1	2	3
Parents	Enfant fumeur	freq
1 Deux	Oui	400
2 Un	Oui	416
3 Aucun	Oui	188
4 Deux	Non	1380
5 Un	Non	1823
6 Aucun	Non	1168

Effet	Niveau Effet	Colonne	Estimat.	Standard Erreur	Wald Stat.	p
Ord.Orig		1	-1,82661	0,078583	540,2950	0,000000
Parents	Deux	2	0,58823	0,096953	36,8105	0,000000
Parents	Un	3	0,34905	0,095539	13,3481	0,000259
Parents	Aucun	4	0,00000			
Echelle			1,00000	0,000000		

On peut déterminer si $\alpha_2 = \alpha_3 = 0$ est une configuration plausible pour les α 's à l'aide de la statistique G2:

The screenshot shows two windows from SPSS. The left window is the 'Résultats' (Results) dialog box, and the right window shows the 'Test Vraisemblance' (Likelihood Ratio Test) results.

The 'Test Vraisemblance Type 1' window displays the following table:

Effet	Degré de Liberté	Log-Vraisbnc	Chi²	p
Ord.Orig	1	-2588,26		
Parents	2	-2569,07	38,36582	0,000000

G2=38.37 et la P-valeur associée indique que non.

Influence de l'inhalation de la fumée et du type de tabac sur le risque du cancer

1. Feuille de données et spécification du modèle de régression logistique

The screenshot displays the GLZ software interface. On the left, a data entry window titled 'Données : vessie.sta (4 var. et 8 obs.)' shows a table with 8 rows and 4 columns: '1 Inhale', '2 Tabac', '3 Cancer', and '4 Effectif'. The data values are as follows:

	1 Inhale	2 Tabac	3 Cancer	4 Effectif
1	O	B	O	267
2	O	A	O	32
3	N	B	O	86
4	N	A	O	23
5	O	B	N	134
6	O	A	N	39
7	N	B	N	119
8	N	A	N	34

On the right, the 'GLZ - Modèles linéaires généraux : vessie.sta' window is open to the 'Avancé' tab. It shows the following settings:

- Variable dépendante : Cancer
- Variable effectifs : --
- Facteurs catégoriels : --
- Prédicteurs continus : --
- Codes réponse : --
- Codes facteurs : --
- Effets inter : --

A dialog box titled 'Sélectionnez la variable dépendante et le prédicteur continu (régresseur) : ' is also visible, showing the selection process for the dependent variable (3 - Cancer) and the continuous predictor (4 - Effectif).

2. Ne pas oublier de décocher 'Sigma-restreint' !

The screenshot shows the 'GLZ - Modèles linéaires généraux : Feuille de données1' window in the 'Avancé' tab. The 'Estimation' section has the following settings:

- Sigma-restreint (unchecked)
- Sans ordonnée origine
- Valeurs départ perso.
- Delta sweep : 1.E-7
- Max. itérations : 100
- Convergence : 7

The 'Construction du modèle' section has the following settings:

- Tous effets
- Pas-à-pas ascendante
- Pas-à-pas descendante
- Inclusion ascendante
- Exclusion descendante
- Recherche exhaustive

On the right, the 'p1, inclusion' and 'p2, exclusion' options are set to 'non'. The 'Max. étapes' and 'Max. sous-ens.' options are also set to 'non'. The 'Score vraisemblance' option is selected, with 'Vraisemblance' and 'Akaike IC' also visible. The 'Validation croisée' option is set to 'non'.

3. Eventuellement changer l'ordre des catégories pour les variables explicatives afin de forcer le choix pour la catégorie de référence :



4. On obtient alors les estimations suivantes :

Cancer - Paramètres estimés (vessie.sta)						
Distribution : BINOMIALE						
Fonction de Liaison : LOGIT						
Effet	Niveau Effet	Colonne	Estimat.	Standard Erreur	Wald Stat.	p
Ord.Orig		1	-0,390866	0,269982	2,095986	0,147686
Inhale	O	2	0,193041	0,360251	0,287135	0,592062
Inhale	N	3	0,000000			
Tabac	B	4	0,066090	0,304830	0,047006	0,828357
Tabac	A	5	0,000000			
Inhale*Tabac	1	6	0,821144	0,401273	4,187535	0,040722
Inhale*Tabac	2	7	0,000000			
Inhale*Tabac	3	8	0,000000			
Inhale*Tabac	4	9	0,000000			
Echelle			1,000000	0,000000		

5. On peut déterminer si $\gamma_{22}=0$ est une configuration plausible à l'aide de la statistique G^2 :

Cancer - Test Vraisemblance Type 1 (vess)				
Distribution : BINOMIALE				
Fonction de Liaison : LOGIT				
Effet	Degré de Liberté	Log-Vraisbnc	Chi ²	p
Ord.Orig.	1	-504,180		
Inhale	1	-488,034	32,29146	0,000000
Tabac	1	-484,281	7,50672	0,006147
Inhale*Tabac	1	-482,209	4,14356	0,041793

$G^2=4.14$ et la P-valeur=0.04 associée inférieure au seuil de 5% suggère que cette configuration n'est pas plausible: l'interaction entre Tabac et Inhalation est dite significative.

Attitude vis-à-vis des homosexuels en fonction de l'âge

1. Feuille des données (homo.sta) et spécification du modèle

The screenshot shows the SPSS 'Données : homo.sta' window with a spreadsheet of data. The data is as follows:

	1	2
	Age	Attitude
1	39	0
2	44	0
3	53	0
4	56	0
5	58	0
6	47	0
7	39	0
8	27	0
9	75	0
10	42	0
11	42	0
12	67	0
13	65	0
14	59	0
15	65	1
16	39	0
17	26	0
18	60	0
19	82	0
20	32	0
21	46	0

Next to it is the 'Sélectionnez la variable dépendante et le prédicteur continu (régresseur) : ' dialog box. The 'Variable dépendante' is set to '2' (Attitude) and the 'Prédicteurs continus' is set to '1' (Age). The 'N'afficher que les variables appropriées' checkbox is unchecked.

2. Forcer la modélisation de la probabilité de 'Attitude=1' :

The screenshot shows the 'GLZ - Modèles linéaires généraux : homo.sta' dialog box. The 'Base' tab is active. The 'Variables' section shows:

- Variable dépendante : Attitude
- Variable effectifs : --
- Facteurs catégoriels : --
- Prédicteurs continus : Age

The 'Codes réponse' section is set to '1 : code de réalisation de l'événement (généralement 1) ; 0 : code de la...'. A small dialog box titled 'Sélectionnez 2 codes pour la réponse bino...' is open, showing 'Attitude:' and the input '1 0'.

3. On obtient alors les estimations suivantes :

Classeur23* - Attitude - Paramètres estimés (homo.sta)

Attitude - Paramètres estimés (homo.sta)
Distribution : BINOMIALE
Fonction de Liaison : LOGIT

Effet	Niveau	Colonne	Estimat	Standard Erreur	Wald Stat.	p
Ord.Orig		1	-4,17399	0,318122	172,1537	0,000000
Age		2	0,02752	0,005852	22,1098	0,000003
Echelle			1,00000	0,000000		

4. Le test G2 confirme que 0 n'est pas une valeur plausible pour beta_1 :

Attitude - Test Vraisemblance Type 1 (horr)				
Distribution : BINOMIALE				
Fonction de Liaison : LOGIT				
Effet	Degré de Liberté	Log-Vraisbnc	Chi ²	p
Ord.Orig.	1	-367,062		
Age	1	-355,732	22,66039	0,000002

Attitude vis-à-vis des homosexuels en fonction de l'âge et du sexe

Nous allons travailler avec l'ensemble des données de l'European Social Science survey pour la Belgique en 2002 (fichier 'ESS2002BE.sta').

1. Chargement de la feuille et construction des variables attitude, âge et sexe

Insertion > Ajouter des variables...

Attitude=1 si freehms égal 5, 0 sinon.

Données : ESS2002BE.sta (583 var. et 1778 obs.)

Ajouter des Variables

Nombre de variables à ajouter : 1

Insérer après la variable : name

Nom : Attitude Type : Double

Code des VM : -9999 Longueur : 8

Format d'affichage

- Général
- Nombre
- Date
- Heure
- Scientifique
- Monétaire
- Pourcentage
- Fraction
- Personnalisé

Description détaillée (étiquette ou formule avec Fonction) :

= (freehms=5)

Formules : utilisez les noms de variables ou v1, v2, ..., v0 est le n° de
Exemples : (a) = mean(v1:v3, sqrt(v7), AGE) (b) = v1+v2; commentaire

Age=2002-yrbrn (yrbrn donne l'année de naissance)

Ajouter des Variables

Nombre de variables à ajouter : 1

Insérer après la variable : Attitude

Nom : Age Type : Double

Code des VM : -9999 Longueur : 8

Format d'affichage

- Général
- Nombre
- Date
- Heure
- Scientifique
- Monétaire
- Pourcentage
- Fraction
- Personnalisé

Description détaillée (étiquette ou formule avec Fonction) :

=2002-yrbrn

Formules : utilisez les noms de variables ou v1, v2, ..., v0 est le n° de
Exemples : (a) = mean(v1:v3, sqrt(v7), AGE) (b) = v1+v2; commentaire

Sexe=gndr

Ajouter des Variables
?
✖

Nombre de variables à ajouter :

Insérer après la variable :

Nom : Type :

Code des VM : Longueur :

Format d'affichage

- Général
- Nombre
- Date
- Heure
- Scientifique
- Monétaire
- Pourcentage
- Fraction
- Personnalisé

Si les valeurs de la nouvelle variable doivent être calculées et que le fichier de données est important, il est plus rapide d'ajouter les variables et de recalculer simultanément leurs valeurs grâce à la commande Formules de Transformation par Lot (du menu Données).

Description détaillée (étiquette ou formule avec) :

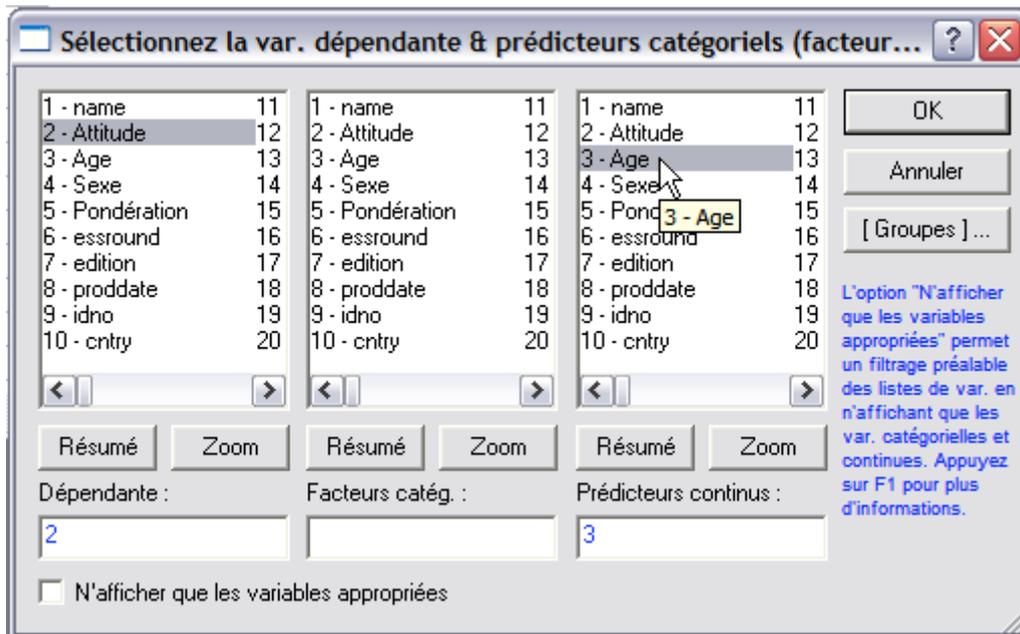
=gndr

Formules : utilisez les noms de variables ou v1, v2, ..., v0 est le n° de
 Exemples : (a) = mean(v1:v3, sqrt(v7), AGE) (b) = v1+v2; commentaire

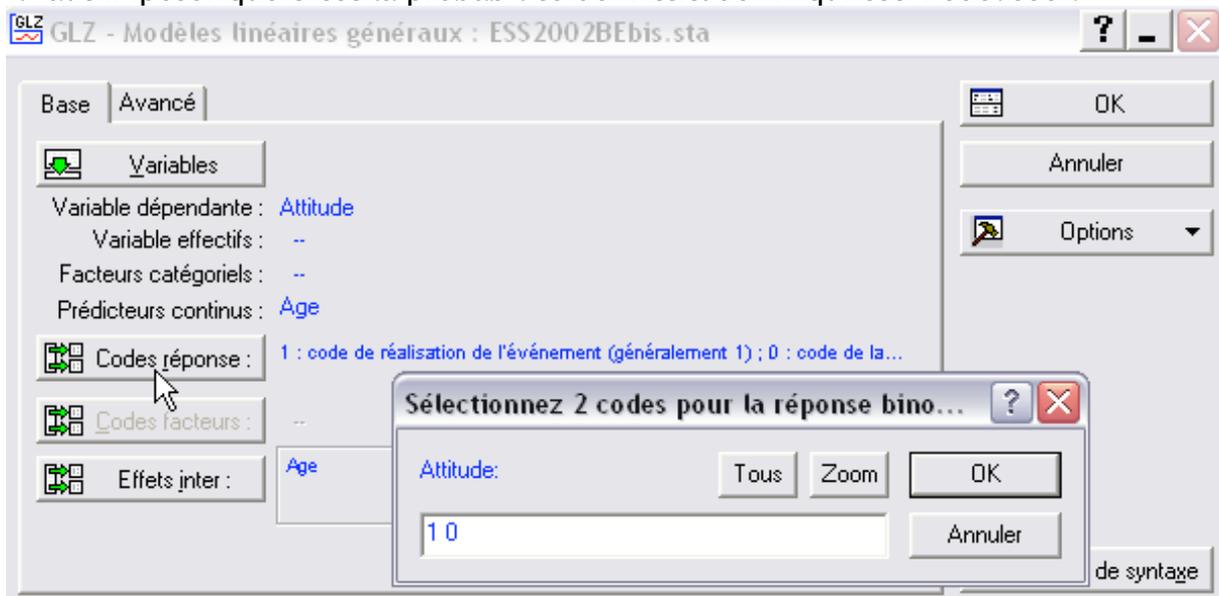
Voici à quoi ressemble désormais la feuille :

Données : ESS2002BEbis.sta (603 var. et 1778 obs.)														
	NAME	Attitude	Age	Sexe	ESSROUND	EDITION	PRODDATE	IDNO	CNTRY	TVTOT	TVPOL	RDTOT	RDPOL	NWSP1
1	ESS2e03	0	39	Male	2	3.0	12.12.2006	####	8,	More th	No time	Less th	No time	No time
2	ESS2e03	0	44	Male	2	3.0	12.12.2006	####	8,	0,5 hou	0,5 hou	Less th	Less th	0,5 hou
3	ESS2e03	0	53	Male	2	3.0	12.12.2006	####	8,	More th				
4	ESS2e03	0	56	Female	2	3.0	12.12.2006	####	8,	More th	0,5 hou	Less th	No time	No time
5	ESS2e03	0	58	Female	2	3.0	12.12.2006	####	8,	More th	More th	More th	No time	Less th
6	ESS2e03	0	47	Female	2	3.0	12.12.2006	####	8,	More th	Less th	0,5 hou	No time	No time
7	ESS2e03	0	39	Male	2	3.0	12.12.2006	####	8,	More th	Less th	0,5 hou	No time	More th
8	ESS2e03	0	27	Male	2	3.0	12.12.2006	####	8,	More th	No time	More th	No time	No time
9	ESS2e03	0	75	Male	2	3.0	12.12.2006	####	8,	More th	0,5 hou	More th	More th	No time
10	ESS2e03	0	42	Male	2	3.0	12.12.2006	####	8,	0,5 hou	Less th	More th	Less th	Less th
11	ESS2e03	0	42	Male	2	3.0	12.12.2006	####	8,	More th	Less th	More th	Less th	0,5 hou
12	ESS2e03	0	67	Male	2	3.0	12.12.2006	####	8,	More th	Less th	No time		More th

2. On peut maintenant ajuster le 1^{er} modèle de régression logisitque :



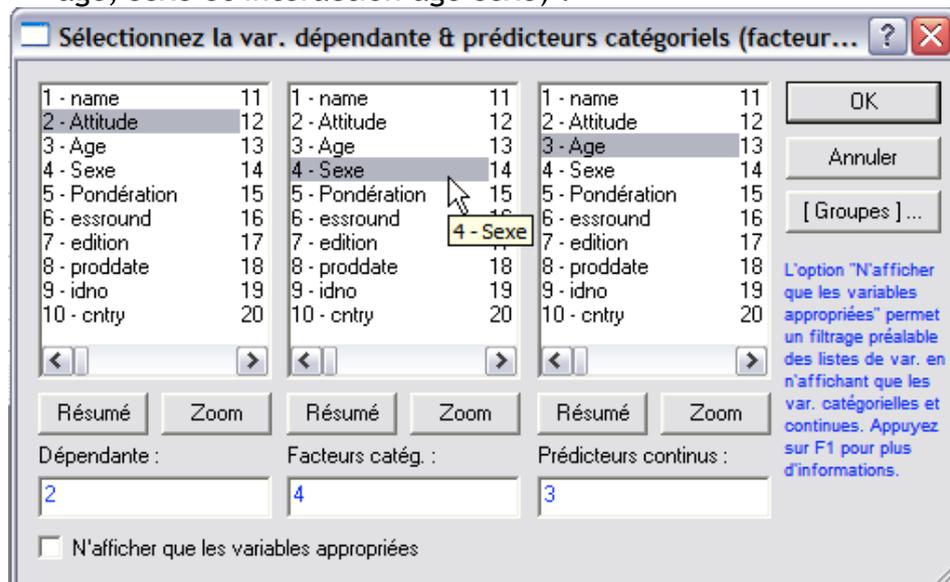
Il faut imposer que c'est la probabilité de 'Attitude=1' qui est modélisée :



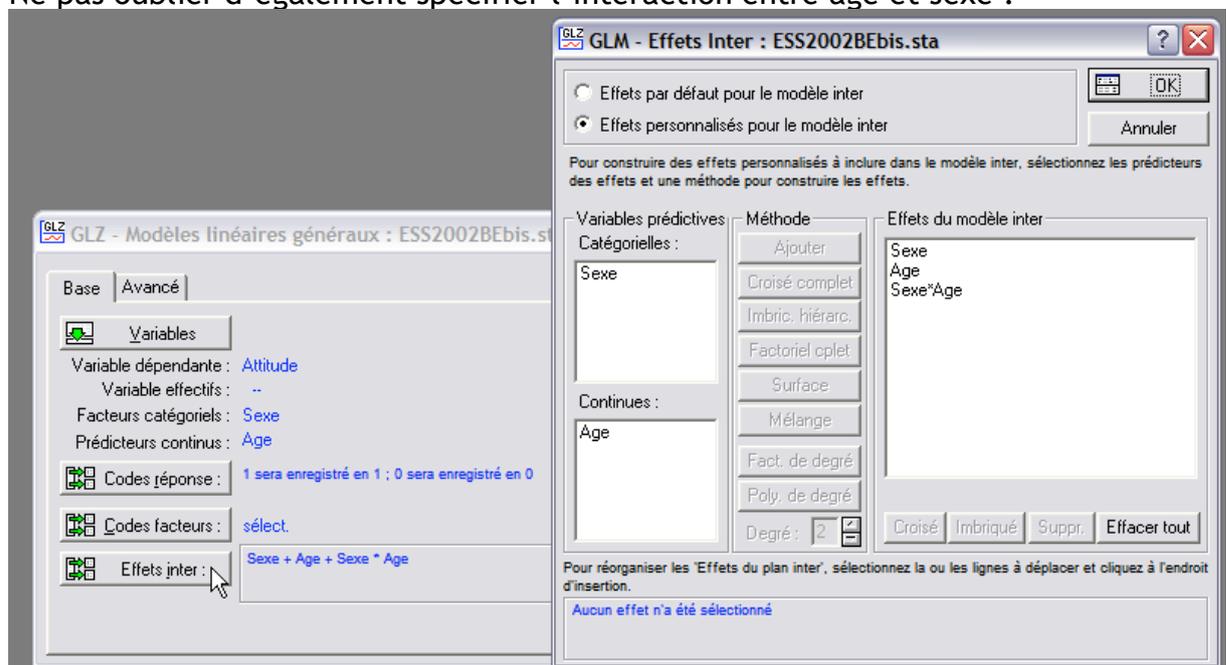
Alors, on obtient les estimations annoncées :

Attitude - Paramètres estimés (ESS2002BEbis.sta)						
Distribution : BINOMIALE						
Fonction de Liaison : LOGIT						
Effet	Niveau Effet	Colonne	Estimat.	Standard Erreur	Wald Stat.	p
Ord.Orig		1	-4,17399	0,318122	172,1537	0,000000
Age		2	0,02752	0,005852	22,1098	0,000003
Echelle			1,00000	0,000000		

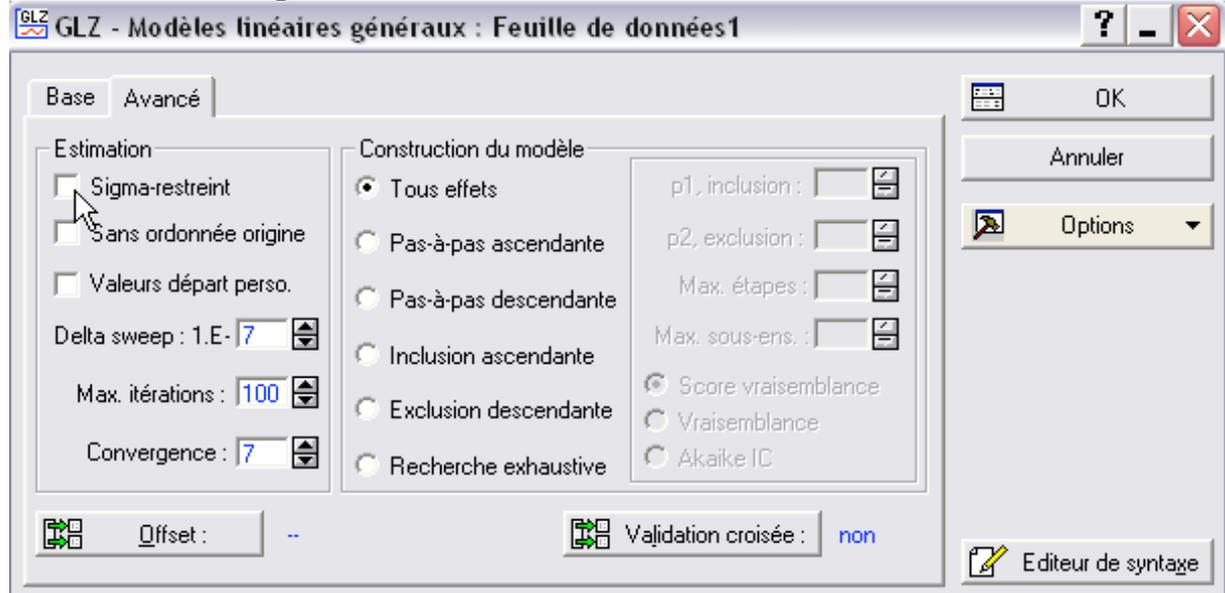
3. On peut maintenant ajuster le 2ème modèle de régression logistique (effets age, sexe et interaction age-sexe) :



Ne pas oublier d'également spécifier l'interaction entre âge et sexe :



et de décocher 'Sigma-restreint'



On obtient ainsi les estimations annoncées :

Attitude - Paramètres estimés (ESS2002BEbis.sta)
 Distribution : BINOMIALE
 Fonction de Liaison : LOGIT

Effet	Niveau Effet	Colonne	Estimat.	Standard Erreur	Wald Stat.	p
Ord.Orig		1	-4,60942	0,486108	89,91388	0,000000
Sexe	Homme	2	0,80923	0,644324	1,57737	0,209140
Sexe	Femme	3	0,00000			
Age		4	0,03445	0,008504	16,41573	0,000051
Sexe*Age	1	5	-0,01317	0,011809	1,24436	0,264633
Sexe*Age	2	6	0,00000			
Echelle			1,00000	0,000000		

Comme précédemment, on peut évaluer la nécessité des différentes variables explicatives via un test de rapport de vraisemblances (cfr. G2):

Le tableau ci-dessous indique qu'il est légitime d'enlever l'interaction Sexe*Age, mais pas Age une fois que l'interaction a déjà été éliminée.

Attitude - Test Vraisemblance Type 1 (ES)				
Distribution : BINOMIALE				
Fonction de Liaison : LOGIT				
Effet	Degré de Liberté	Log-Vraisbnc	Chi ²	p
Ord.Orig.	1	-367,062		
Sexe	1	-366,993	0,13704	0,711237
Age	1	-355,537	22,91228	0,000002
Sexe*Age	1	-354,912	1,25118	0,263327

Le tableau ci-dessous indique qu'il est légitime d'enlever l'interaction Sexe*Age. De même, il est légitime d'enlever Sexe une fois que l'interaction Sexe*Age est éliminée. Cependant, la variable Age doit être maintenue.

Attitude - Test Vraisemblance Type 1 (ES)				
Distribution : BINOMIALE				
Fonction de Liaison : LOGIT				
Effet	Degré de Liberté	Log-Vraisbnc	Chi ²	p
Ord.Orig.	1	-367,062		
Age	1	-355,732	22,66039	0,000002
Sexe	1	-355,537	0,38893	0,532861
Sexe*Age	1	-354,912	1,25118	0,263327