

EXEMPLE D'UTILISATION DE LA PROCEDURE CATMOD

ALIMENTATION DES ALLIGATORS

Le tableau 1 provient d'une étude concernant les facteurs influençant le choix de la nourriture principale d'alligators. L'échantillon est formé de 219 alligators capturés dans quatre lacs de Floride. La variable à modéliser est la nourriture principale, en volume, trouvée dans l'estomac des alligators. Cette variable (Choix) a cinq catégories : poissons, invertébrés, autres, oiseaux, reptiles.

Les variables explicatives, outre le lac, sont le sexe des alligators (Genre) : male ou femelle, et leur taille codée en deux catégories :
 taille 1 ($\leq 2,3$ mètres)
 taille 2 ($>2,3$ mètres).

On trouve dans les cinq colonnes de droite le nombre d'alligators qui se sont nourris principalement de poissons, d'invertébrés, d'oiseaux, de reptiles ou d'autres nourritures.

Lac	Genre	Taille	Poissons	Invertébrés	Autres	Oiseaux (Birds)	Reptiles
Hancock	Male	1	7	1	5	0	0
		2	4	0	2	1	0
	Femelle	1	16	3	3	2	2
		2	3	0	3	2	1
Oklawaha	Male	1	2	2	1	0	0
		2	13	7	0	0	6
	Femelle	1	3	9	2	0	1
		2	0	1	0	1	0
Trafford	Male	1	3	7	1	0	1
		2	8	6	5	3	6
	Femelle	1	2	4	4	1	1
		2	0	1	0	0	0
George	Male	1	13	10	2	2	0
		2	9	0	2	1	0
	Femelle	1	3	9	1	0	1
		2	8	1	1	0	0

TABLEAU 1

- 1) L'objectif de l'étude est de construire un modèle explicatif de la variable choix à l'aide des variables lac, genre, taille. Dans ce but on utilise la procédure CATMOD. Indiquer les cinq options disponibles pour les fonctions de réponse. Lesquelles sont utilisables pour ce type de données ?
- 2) Comment doit on présenter les données pour le programme CATMOD ? On indiquera en particulier le nombre de lignes.
- 3) Indiquer ce que l'on entend par fonction de réponse logit généralisé. En utilisant les notations du cours, on écrira précisément les différents termes. On démontrera comment on en déduit les probabilités associées aux différentes modalités de la variable à expliquer Y.
- 4) Les résultats obtenus avec l'option par défaut de CATMOD pour le modèle avec uniquement les effets principaux : lac, genre, taille sont fournis dans le tableau 2. Que pensez vous de ce modèle ? Tester l'apport de chaque variable.
- 5) Les résultats concernant le modèle simplifié (effets lac et taille) sont présentés dans le tableau 3. Quelle est l'utilité dans la procédure CATMOD de l'instruction : « **population** lac genre taille; ». Pourquoi n'était-elle pas nécessaire dans le modèle avec les trois effets principaux lac genre taille ? (tableau 2)
- 6) Commenter les résultats du tableau 3.
- 7) A partir du modèle simplifié, calculer pour les males de taille 1 du lac Hancock, les probabilités estimées des choix de nourriture poissons, invertébrés. Indiquer aussi les probabilités estimées des choix de nourriture poissons, invertébrés pour les femelles de taille 1 du lac Hancock. (compléter le tableau 4).
- 8) Le tableau 5 présente les sorties d'une modélisation à l'aide de la fonction identité. Un message d'erreur apparaît. Comment peut-on l'expliquer ? Comment peut-on éviter ce type de problème dans SAS ?
- 9) Le tableau 6 présente les résultats du modèle utilisant la fonction identité, prenant en compte les effets principaux lac genre taille. Que pensez-vous de ce modèle ? Que peut-on proposer ?

ANNEXE : ALIMENTATION DES ALLIGATORS

TABLEAU 2 : Procédure CATMOD ; Modèle avec les effets principaux lac, genre, taille.

```
proc catmod data=alligat1 order=data;
weight freq;
model choix= lac genre taille;
run;
```

Data Summary

Response	choix	Response Levels	5
Weight Variable	freq	Populations	16
Data Set	ALLIGAT1	Total Frequency	219
Frequency Missing	0	Observations	56

Population Profiles

Sample	lac	genre	taille	Sample Size
1	H	M	1	13
2	H	M	2	7
3	H	F	1	26
4	H	F	2	9
5	O	M	1	5
6	O	M	2	26
7	O	F	1	15
8	O	F	2	2
9	T	M	1	12
10	T	M	2	28
11	T	F	1	12
12	T	F	2	1
13	G	M	1	27
14	G	M	2	12
15	G	F	1	14
16	G	F	2	10

Response Profiles

Response	choix
1	P
2	I
3	A
4	B
5	R

Maximum Likelihood Analysis of Variance

Source	DF	Chi-Square	Pr > ChiSq
Intercept	4	62.04	<.0001
lac	12	36.23	0.0003
genre	4	2.18	0.7018
taille	4	15.89	0.0032
Likelihood Ratio	40	50.26	0.1282

Analysis of Maximum Likelihood Estimates

Effect	Parameter	Estimate	Standard Error	Chi-Square	Pr > ChiSq
Intercept	1	1.7711	0.3440	26.50	<.0001
	2	1.1126	0.3708	9.00	0.0027
	3	0.6564	0.3863	2.89	0.0893
	4	-0.2836	0.4762	0.35	0.5515
lac	5	0.5507	0.5777	0.91	0.3405
	6	-1.3019	0.6768	3.70	0.0544
	7	0.7297	0.6270	1.35	0.2445
	8	0.8105	0.7441	1.19	0.2760
	9	-0.8501	0.4735	3.22	0.0726
	10	-0.00901	0.4962	0.00	0.9855
	11	-1.4116	0.6289	5.04	0.0248
	12	-1.7159	0.8988	3.64	0.0562
	13	-1.3809	0.4843	8.13	0.0044
	14	-0.2972	0.4953	0.36	0.5485
	15	-0.4107	0.5294	0.60	0.4379
	16	-0.4593	0.6645	0.48	0.4894
genre	17	0.3138	0.3426	0.84	0.3598
	18	0.0823	0.3510	0.05	0.8146
	19	0.1875	0.3777	0.25	0.6196
taille	20	0.0106	0.4544	0.00	0.9815
	21	0.2785	0.3233	0.74	0.3890
	22	0.9466	0.3386	7.82	0.0052
	23	0.4238	0.3573	1.41	0.2356
	24	-0.0866	0.4277	0.04	0.8396

TABLEAU 3 : MODELE SIMPLIFIE, EFFETS LAC ET TAILLE

```
proc catmod data=alligat1 order=data;
weight freq;
population lac genre taille;
model choix= lac taille / freq prob predict ;
run;
```

Maximum Likelihood Analysis of Variance

Source	DF	Chi-Square	Pr > ChiSq
Intercept	4	70.39	<.0001
lac	12	35.49	0.0004
taille	4	18.76	0.0009
Likelihood Ratio	44	52.48	0.1784

Analysis of Maximum Likelihood Estimates

Effect	Parameter	Estimate	Standard Error	Chi-Square	Pr > ChiSq
Intercept	1	1.8309	0.3398	29.03	<.0001
	2	1.1112	0.3682	9.11	0.0025
	3	0.6795	0.3818	3.17	0.0751
	4	-0.2950	0.4738	0.39	0.5334
lac	5	0.4164	0.5589	0.56	0.4562
	6	-1.3421	0.6641	4.08	0.0433
	7	0.6556	0.6071	1.17	0.2802
	8	0.8291	0.7140	1.35	0.2455
	9	-0.7996	0.4710	2.88	0.0896
	10	0.0374	0.4962	0.01	0.9400
	11	-1.3810	0.6279	4.84	0.0278
	12	-1.7353	0.8953	3.76	0.0526
	13	-1.2760	0.4677	7.44	0.0064
	14	-0.2543	0.4811	0.28	0.5971
	15	-0.3467	0.5130	0.46	0.4991
	16	-0.4707	0.6400	0.54	0.4621
taille	17	0.1756	0.2900	0.37	0.5448
	18	0.9047	0.3014	9.01	0.0027
	19	0.3414	0.3257	1.10	0.2945
	20	-0.1397	0.4031	0.12	0.7289

TABLEAU 4 : EXTRAIT DU TABLEAU DES PREVISIONS MODELE SIMPLIFIE.

Maximum Likelihood Predicted Values for Probabilities

lac	genre	taille	choix	-----Observed-----		-----Predicted-----		Residual
				Probability	Standard Error	Probability	Standard Error	
H	M	1	P	0.5385	0.1383	???	0.0721	???
			I	0.0769	0.0739	???	0.0443	???
			A	0.3846	0.1349	0.2537	0.0637	0.1309
			B	0	0	0.0704	0.0356	-0.07
			R	0	0	0.0475	0.0285	-0.047
H	M	2	P	0.5714	0.187	0.5702	0.0912	0.0012
			I	0	0	0.0231	0.014	-0.023
			A	0.2857	0.1707	0.194	0.0709	0.0917
			B	0.1429	0.1323	0.1409	0.0716	0.002
			R	0	0	0.0718	0.046	-0.072
H	F	1	P	0.6154	0.0954	???	0.0721	???
			I	0.1154	0.0627	???	0.0443	0.0223
			A	0.1154	0.0627	0.2537	0.0637	-0.138
			B	0.0769	0.0523	0.0704	0.0356	0.0065
			R	0.0769	0.0523	0.0475	0.0285	0.0295
H	F	2	P	0.3333	0.1571	0.5702	0.0912	-0.237
			I	0	0	0.0231	0.014	-0.023
			A	0.3333	0.1571	0.194	0.0709	0.1393
			B	0.2222	0.1386	0.1409	0.0716	0.0813
			R	0.1111	0.1048	0.0718	0.046	0.0393

TABLEAU 5 : MODELE FONCTION IDENTITE

```
proc catmod data=alligat1 order=data;
weight freq;
response marginals;
model choix= lac genre taille;
run;
```

The CATMOD Procedure

Data Summary

Response	choix	Response Levels	5
Weight Variable	freq	Populations	16
Data Set	ALLIGAT1	Total Frequency	219
Frequency Missing	0	Observations	56

Population Profiles

Sample	lac	genre	taille	Sample Size
1	H	M	1	13
2	H	M	2	7
3	H	F	1	26
4	H	F	2	9
5	O	M	1	5
6	O	M	2	26
7	O	F	1	15
8	O	F	2	2
9	T	M	1	12
10	T	M	2	28
11	T	F	1	12
12	T	F	2	1
13	G	M	1	27
14	G	M	2	12
15	G	F	1	14
16	G	F	2	10

Response choix

1	P
2	I
3	A
4	B
5	R

ERROR: The response functions are linearly dependent since the number of functions per population, 4, is greater than or equal to the number of response levels, 1, in population 12.

TABLEAU 6 : MODELE FONCTION IDENTITE effectifs modifiés .

```
proc catmod data=alligat1 order=data;
weight freq;
response marginals;
population lac genre taille;
model choix= lac genre taille/ addcell = 0.1 ;
run;
```

Analysis of Variance

Source	DF	Chi-Square	Pr > ChiSq
Intercept	4	3232.62	<.0001
lac	12	124.89	<.0001
genre	4	11.90	0.0181
taille	4	24.33	<.0001
Residual	40	63.04	0.0115

Analysis of Weighted Least Squares Estimates

Effect	Parameter	Estimate	Standard Error	Chi-Square	Pr > ChiSq
Intercept	1	0.4631	0.0272	289.25	<.0001
	2	0.2881	0.0249	133.89	<.0001
	3	0.1400	0.0206	46.07	<.0001
lac	4	0.0316	0.0111	8.15	0.0043
	5	0.1676	0.0489	11.75	0.0006
	6	-0.2047	0.0297	47.66	<.0001
	7	0.0607	0.0437	1.93	0.1649
	8	0.0169	0.0190	0.79	0.3747
	9	-0.1168	0.0492	5.65	0.0175
	10	0.2349	0.0476	24.36	<.0001
	11	-0.1180	0.0306	14.84	0.0001
	12	-0.0253	0.0137	3.39	0.0656
	13	-0.2298	0.0470	23.92	<.0001
	14	0.0890	0.0469	3.60	0.0579
genre	15	0.0526	0.0392	1.80	0.1794
	16	0.0179	0.0197	0.83	0.3629
	17	0.0884	0.0260	11.58	0.0007
	18	-0.0352	0.0185	3.60	0.0576
	19	-0.0303	0.0214	2.02	0.1555
taille	20	-0.0105	0.0112	0.87	0.3514
	21	-0.0580	0.0270	4.62	0.0316
	22	0.0958	0.0198	23.48	<.0001
	23	-0.0168	0.0223	0.56	0.4523
	24	-0.0132	0.0106	1.56	0.2121