

EXEMPLE D'UTILISATION DE CATMOD :

ACCIDENTS

Les données de cet exemple proviennent d'une étude concernant la gravité d'accidents de voitures en fonction de la présence ou de l'utilisation sur le véhicule de divers équipements. On note pour chaque collision de type frontal si le conducteur est gravement blessé ou non.

Les variables en présence sont les suivantes :

Présence d'un système d'absorption de choc au niveau du volant

Usage de la ceinture de sécurité

Vitesse (en miles par heure) avant le choc

Automobile's Energy-Absorbing steering status	Seatbelt usage	Speed of car prior to impact	Driver injury status		
			Serious	Not serious	Total
Absent	No	0-29	31	1419	1450
Absent	No	30-49	191	2004	2195
Absent	No	50+	216	1030	1246
Absent	Yes	0-29	6	255	261
Absent	Yes	30-49	14	339	353
Absent	Yes	50+	19	200	219
Present	No	0-29	22	652	674
Present	No	30-49	57	928	985
Present	No	50+	108	515	623
Present	Yes	0-29	4	199	203
Present	Yes	30-49	11	265	276
Present	Yes	50+	20	157	177

Questions

- 1) Construire le modèle saturé.
- 2) Construire un modèle satisfaisant en éliminant progressivement les termes non significatifs.
- 3) Construire un modèle équivalent au modèle simplifié retenu en imbriquant Speed (avec ses différents niveaux) dans Steering et Seatbelt.
- 4) Eliminer les termes non significatifs. Ecrire l'équation du modèle retenu. Interpréter les résultats du point de vue de la sécurité.
- 5) Comparer graphiquement proportions observées et estimées.

PROGRAMME SAS

```
Data highway;
Input steering $ seatbelt $ speed $ status $ count;
cards;
absent no 0-29 serious 31
absent no 0-29 not 1419
absent no 30-49 serious 191
absent no 30-49 not 2004
absent no 50+ serious 216
absent no 50+ not 1030
absent yes 0-29 serious 6
absent yes 0-29 not 255
absent yes 30-49 serious 14
absent yes 30-49 not 339
absent yes 50+ serious 19
absent yes 50+ not 200
present no 0-29 serious 22
present no 0-29 not 652
present no 30-49 serious 57
present no 30-49 not 928
present no 50+ serious 108
present no 50+ not 515
present yes 0-29 serious 4
present yes 0-29 not 199
present yes 30-49 serious 11
present yes 30-49 not 265
present yes 50+ serious 20
present yes 50+ not 157 ;

proc catmod data=highway order= data;
weight count;
population steering seatbelt speed;
response marginals;
model status = steering | seatbelt | speed / freq;

proc catmod data=highway order= data;
weight count;
population steering seatbelt speed;
response marginals;
model status = steering seatbelt speed
steering*seatbelt steering*speed seatbelt*speed/ noprofile;

proc catmod data=highway order= data;
weight count;
population steering seatbelt speed;
response marginals;
model status = steering seatbelt speed
steering*speed seatbelt*speed/ noprofile;
```

```

proc catmod data=highway order= data;
weight count;
population steering seatbelt speed;
response marginals;
model status = speed
    steering(speed='0-29')
    steering(speed='30-49')
    steering(speed='50+')
    seatbelt(speed='0-29')
    seatbelt(speed='30-49')
    seatbelt(speed='50+') / noprofile;

```

```

proc catmod data=highway order= data;
weight count;
population steering seatbelt speed;
response marginals;
model status = speed
    steering(speed='30-49')
    steering(speed='50+')
    seatbelt(speed='30-49')
    seatbelt(speed='50+') / noprofile;

```

```

proc catmod data=highway order= data;
weight count;
population steering seatbelt speed;
response marginals / out =a;
model status = speed
    steering(speed='30-49')
    seatbelt(speed='30-49')
    seatbelt(speed='50+') / pred=prob noprofile;

```

```

proc print data=a;
data b; set a;i=_n_;
proc plot data=b;
plot _obs_* _pred_ = ' ' $ i
    / box haxis = by .04 vaxis = by .04 overlay;
run;

```

TABLEAU 1: MODELE SATURE

CATMOD PROCEDURE

Response: STATUS	Response Levels (R)= 2
Weight Variable: COUNT	Populations (S)= 12
Data Set: HIGHWAY	Total Frequency (N)= 8662
Frequency Missing: 0	Observations (Obs)= 24

POPULATION PROFILES

Sample	STEERING	SEATBELT	SPEED	Sample Size
1	absent	no	0-29	1450
2	absent	no	30-49	2195
3	absent	no	50+	1246
4	absent	yes	0-29	261
5	absent	yes	30-49	353
6	absent	yes	50+	219
7	present	no	0-29	674
8	present	no	30-49	985
9	present	no	50+	623
10	present	yes	0-29	203
11	present	yes	30-49	276
12	present	yes	50+	177

RESPONSE PROFILES

Response	STATUS
1	serious
2	not

RESPONSE FREQUENCIES

Sample	Response Number 1	Response Number 2
1	31	1419
2	191	2004
3	216	1030
4	6	255
5	14	339
6	19	200
7	22	652
8	57	928
9	108	515
10	4	199
11	11	265
12	20	157

Sample	Response Function	DESIGN MATRIX											
		1	2	3	4	5	6	7	8	9	10	11	12
1	0.02138	1	1	1	1	1	0	1	0	1	0	1	0
2	0.08702	1	1	1	1	0	1	0	1	0	1	0	1
3	0.17335	1	1	1	1	-1	-1	-1	-1	-1	-1	-1	-1
4	0.02299	1	1	-1	-1	1	0	1	0	-1	0	-1	0
5	0.03966	1	1	-1	-1	0	1	0	1	0	-1	0	-1
6	0.08676	1	1	-1	-1	-1	-1	-1	-1	1	1	1	1
7	0.03264	1	-1	1	-1	1	0	-1	0	1	0	-1	0
8	0.05787	1	-1	1	-1	0	1	0	-1	0	1	0	-1
9	0.17335	1	-1	1	-1	-1	-1	1	1	-1	-1	1	1
10	0.01970	1	-1	-1	1	1	0	-1	0	-1	0	1	0
11	0.03986	1	-1	-1	1	0	1	0	-1	0	-1	0	1
12	0.11299	1	-1	-1	1	-1	-1	1	1	1	1	-1	-1

ANALYSIS-OF-VARIANCE TABLE

Source	DF	Chi-Square	Prob
INTERCEPT	1	406.06	0.0000
STEERING	1	0.01	0.9027
SEATBELT	1	26.99	0.0000
STEERING*SEATBELT	1	0.91	0.3406
SPEED	2	139.20	0.0000
STEERING*SPEED	2	3.15	0.2069
SEATBELT*SPEED	2	14.07	0.0009
STEERING*SEATBELT*SPEED	2	3.67	0.1593
RESIDUAL	0	.	.

ANALYSIS OF WEIGHTED-LEAST-SQUARES ESTIMATES

Effect	Parameter	Estimate	Standard Error	Chi-Square	Prob
INTERCEPT	1	0.0723	0.00359	406.06	0.0000
STEERING	2	-0.00044	0.00359	0.01	0.9027
SEATBELT	3	0.0186	0.00359	26.99	0.0000
STEERING*SEATBELT	4	0.00342	0.00359	0.91	0.3406
SPEED	5	-0.0481	0.00423	129.18	0.0000
	6	-0.0162	0.00446	13.17	0.0003
STEERING*SPEED	7	-0.00156	0.00423	0.14	0.7132
	8	0.00768	0.00446	2.96	0.0854
SEATBELT*SPEED	9	-0.0158	0.00423	13.94	0.0002
	10	-0.00230	0.00446	0.26	0.6070
STEERING*SEATBELT*SPEED	11	-0.00706	0.00423	2.78	0.0956
	12	0.00392	0.00446	0.77	0.3802

TABLEAU 2: MODELE SANS INTERACTION D'ORDRE 3

CATMOD PROCEDURE

Response: STATUS
 Weight Variable: COUNT
 Data Set: HIGHWAY
 Frequency Missing: 0

Response Levels (R)= 2
 Populations (S)= 12
 Total Frequency (N)= 8662
 Observations (Obs)= 24

Sample	Response Function	DESIGN MATRIX									
		1	2	3	4	5	6	7	8	9	10
1	0.02138	1	1	1	1	0	1	1	0	1	0
2	0.08702	1	1	1	0	1	1	0	1	0	1
3	0.17335	1	1	1	-1	-1	1	-1	-1	-1	-1
4	0.02299	1	1	-1	1	0	-1	1	0	-1	0
5	0.03966	1	1	-1	0	1	-1	0	1	0	-1
6	0.08676	1	1	-1	-1	-1	-1	-1	-1	1	1
7	0.03264	1	-1	1	1	0	-1	-1	0	1	0
8	0.05787	1	-1	1	0	1	-1	0	-1	0	1
9	0.17335	1	-1	1	-1	-1	-1	1	1	-1	-1
10	0.01970	1	-1	-1	1	0	1	-1	0	-1	0
11	0.03986	1	-1	-1	0	1	1	0	-1	0	-1
12	0.11299	1	-1	-1	-1	-1	1	1	1	1	1

ANALYSIS-OF-VARIANCE TABLE

Source	DF	Chi-Square	Prob
INTERCEPT	1	404.08	0.0000
STEERING	1	0.01	0.9132
SEATBELT	1	29.04	0.0000
SPEED	2	140.13	0.0000
STEERING*SEATBELT	1	0.29	0.5930
STEERING*SPEED	2	7.87	0.0196
SEATBELT*SPEED	2	17.19	0.0002
RESIDUAL	2	3.67	0.1593

ANALYSIS OF WEIGHTED-LEAST-SQUARES ESTIMATES

Effect	Parameter	Estimate	Standard Error	Chi-Square	Prob
INTERCEPT	1	0.0719	0.00358	404.08	0.0000
STEERING	2	0.000376	0.00345	0.01	0.9132
SEATBELT	3	0.0191	0.00354	29.04	0.0000
SPEED	4	-0.0483	0.00423	130.18	0.0000
	5	-0.0160	0.00446	12.97	0.0003
STEERING*SEATBELT	6	0.00151	0.00282	0.29	0.5930
STEERING*SPEED	7	-0.00491	0.00373	1.73	0.1880
	8	0.00953	0.00396	5.80	0.0160
SEATBELT*SPEED	9	-0.0171	0.00414	17.08	0.0000
	10	-0.00185	0.00437	0.18	0.6721

TABLEAU 3: MODELE SANS INTERACTION D'ORDRE 3 et SANS INTERACTION

STEERING*SEATBELT

CATMOD PROCEDURE

Response: STATUS
 Weight Variable: COUNT
 Data Set: HIGHWAY
 Frequency Missing: 0

Response Levels (R)= 2
 Populations (S)= 12
 Total Frequency (N)= 8662
 Observations (Obs)= 24

Sample	Response Function	DESIGN MATRIX								
		1	2	3	4	5	6	7	8	9
1	0.02138	1	1	1	1	0	1	0	1	0
2	0.08702	1	1	1	0	1	0	1	0	1
3	0.17335	1	1	1	-1	-1	-1	-1	-1	-1
4	0.02299	1	1	-1	1	0	1	0	-1	0
5	0.03966	1	1	-1	0	1	0	1	0	-1
6	0.08676	1	1	-1	-1	-1	-1	-1	1	1
7	0.03264	1	-1	1	1	0	-1	0	1	0
8	0.05787	1	-1	1	0	1	0	-1	0	1
9	0.17335	1	-1	1	-1	-1	1	1	-1	-1
10	0.01970	1	-1	-1	1	0	-1	0	-1	0
11	0.03986	1	-1	-1	0	1	0	-1	0	-1
12	0.11299	1	-1	-1	-1	-1	1	1	1	1

ANALYSIS-OF-VARIANCE TABLE

Source	DF	Chi-Square	Prob
INTERCEPT	1	404.04	0.0000
STEERING	1	0.12	0.7331
SEATBELT	1	30.62	0.0000
SPEED	2	139.86	0.0000
STEERING*SPEED	2	7.82	0.0201
SEATBELT*SPEED	2	17.26	0.0002
RESIDUAL	3	3.96	0.2658

ANALYSIS OF WEIGHTED-LEAST-SQUARES ESTIMATES

Effect	Parameter	Estimate	Standard Error	Chi-Square	Prob
INTERCEPT	1	0.0719	0.00358	404.04	0.0000
STEERING	2	0.00108	0.00318	0.12	0.7331
SEATBELT	3	0.0193	0.00350	30.62	0.0000
SPEED	4	-0.0481	0.00422	129.95	0.0000
	5	-0.0161	0.00445	13.05	0.0003
STEERING*SPEED	6	-0.00488	0.00373	1.71	0.1913
	7	0.00952	0.00396	5.78	0.0162
SEATBELT*SPEED	8	-0.0171	0.00414	17.14	0.0000
	9	-0.00191	0.00437	0.19	0.6613

TABLEAU 4: MODELE IMBRIQUANT SPEED DANS STEERING et SEATBELT

CATMOD PROCEDURE

Response: STATUS
 Weight Variable: COUNT
 Data Set: HIGHWAY
 Frequency Missing: 0

Response Levels (R)= 2
 Populations (S)= 12
 Total Frequency (N)= 8662
 Observations (Obs)= 24

Sample	Response Function	DESIGN MATRIX								
		1	2	3	4	5	6	7	8	9
1	0.02138	1	1	0	1	0	0	1	0	0
2	0.08702	1	0	1	0	1	0	0	1	0
3	0.17335	1	-1	-1	0	0	1	0	0	1
4	0.02299	1	1	0	1	0	0	-1	0	0
5	0.03966	1	0	1	0	1	0	0	-1	0
6	0.08676	1	-1	-1	0	0	1	0	0	-1
7	0.03264	1	1	0	-1	0	0	1	0	0
8	0.05787	1	0	1	0	-1	0	0	1	0
9	0.17335	1	-1	-1	0	0	-1	0	0	1
10	0.01970	1	1	0	-1	0	0	-1	0	0
11	0.03986	1	0	1	0	-1	0	0	-1	0
12	0.11299	1	-1	-1	0	0	-1	0	0	-1

ANALYSIS-OF-VARIANCE TABLE

Source	DF	Chi-Square	Prob
INTERCEPT	1	404.04	0.0000
SPEED	2	139.86	0.0000
STEERING(SPEED=0-29)	1	1.26	0.2623
STEERING(SPEED=30-49)	1	6.73	0.0095
STEERING(SPEED=50+)	1	0.20	0.6539
SEATBELT(SPEED=0-29)	1	0.33	0.5646
SEATBELT(SPEED=30-49)	1	14.69	0.0001
SEATBELT(SPEED=50+)	1	19.76	0.0000
RESIDUAL	3	3.96	0.2658

ANALYSIS OF WEIGHTED-LEAST-SQUARES ESTIMATES

Effect	Parameter	Estimate	Standard Error	Chi-Square	Prob
INTERCEPT	1	0.0719	0.00358	404.04	0.0000
SPEED	2	-0.0481	0.00422	129.95	0.0000
	3	-0.0161	0.00445	13.05	0.0003
STEERING(SPEED=0-29)	4	-0.00379	0.00338	1.26	0.2623
STEERING(SPEED=30-49)	5	0.0106	0.00409	6.73	0.0095
STEERING(SPEED=50+)	6	-0.00356	0.00793	0.20	0.6539
SEATBELT(SPEED=0-29)	7	0.00221	0.00384	0.33	0.5646
SEATBELT(SPEED=30-49)	8	0.0174	0.00455	14.69	0.0001
SEATBELT(SPEED=50+)	9	0.0384	0.00863	19.76	0.0000

TABLEAU 5: MODELE IMBRIQUANT SPEED DANS STEERING et SEATBELT APRES ELIMINATION DE TERMES NON SIGNIFICATIFS (1)

CATMOD PROCEDURE

Response: STATUS
 Weight Variable: COUNT
 Data Set: HIGHWAY
 Frequency Missing: 0

Response Levels (R)= 2
 Populations (S)= 12
 Total Frequency (N)= 8662
 Observations (Obs)= 24

Sample	Response Function	DESIGN MATRIX						
		1	2	3	4	5	6	7
1	0.02138	1	1	0	0	0	0	0
2	0.08702	1	0	1	1	0	1	0
3	0.17335	1	-1	-1	0	1	0	1
4	0.02299	1	1	0	0	0	0	0
5	0.03966	1	0	1	1	0	-1	0
6	0.08676	1	-1	-1	0	1	0	-1
7	0.03264	1	1	0	0	0	0	0
8	0.05787	1	0	1	-1	0	1	0
9	0.17335	1	-1	-1	0	-1	0	1
10	0.01970	1	1	0	0	0	0	0
11	0.03986	1	0	1	-1	0	-1	0
12	0.11299	1	-1	-1	0	-1	0	-1

ANALYSIS-OF-VARIANCE TABLE

Source	DF	Chi-Square	Prob
INTERCEPT	1	425.58	0.0000
SPEED	2	159.24	0.0000
STEERING(SPEED=30-49)	1	6.73	0.0095
STEERING(SPEED=50+)	1	0.20	0.6539
SEATBELT(SPEED=30-49)	1	14.69	0.0001
SEATBELT(SPEED=50+)	1	19.76	0.0000
RESIDUAL	5	5.34	0.3761

ANALYSIS OF WEIGHTED-LEAST-SQUARES ESTIMATES

Effect	Parameter	Estimate	Standard Error	Chi-Square	Prob
INTERCEPT	1	0.0718	0.00348	425.58	0.0000
SPEED	2	-0.0483	0.00388	154.80	0.0000
	3	-0.0160	0.00438	13.35	0.0003
STEERING(SPEED=30-49)	4	0.0106	0.00409	6.73	0.0095
STEERING(SPEED=50+)	5	-0.00356	0.00793	0.20	0.6539
SEATBELT(SPEED=30-49)	6	0.0174	0.00455	14.69	0.0001
SEATBELT(SPEED=50+)	7	0.0384	0.00863	19.76	0.0000

TABLEAU 6: MODELE IMBRIQUANT SPEED DANS STEERING et SEATBELT APRES ELIMINATION DE TERMES NON SIGNIFICATIFS (2)

CATMOD PROCEDURE

Response: STATUS
 Weight Variable: COUNT
 Data Set: HIGHWAY
 Frequency Missing: 0

Response Levels (R)= 2
 Populations (S)= 12
 Total Frequency (N)= 8662
 Observations (Obs)= 24

Sample	Response Function	DESIGN MATRIX					
		1	2	3	4	5	6
1	0.02138	1	1	0	0	0	0
2	0.08702	1	0	1	1	1	0
3	0.17335	1	-1	-1	0	0	1
4	0.02299	1	1	0	0	0	0
5	0.03966	1	0	1	1	-1	0
6	0.08676	1	-1	-1	0	0	-1
7	0.03264	1	1	0	0	0	0
8	0.05787	1	0	1	-1	1	0
9	0.17335	1	-1	-1	0	0	1
10	0.01970	1	1	0	0	0	0
11	0.03986	1	0	1	-1	-1	0
12	0.11299	1	-1	-1	0	0	-1

ANALYSIS-OF-VARIANCE TABLE

Source	DF	Chi-Square	Prob
INTERCEPT	1	441.18	0.0000
SPEED	2	164.17	0.0000
STEERING(SPEED=30-49)	1	6.73	0.0095
SEATBELT(SPEED=30-49)	1	14.69	0.0001
SEATBELT(SPEED=50+)	1	19.61	0.0000
RESIDUAL	6	5.54	0.4769

ANALYSIS OF WEIGHTED-LEAST-SQUARES ESTIMATES

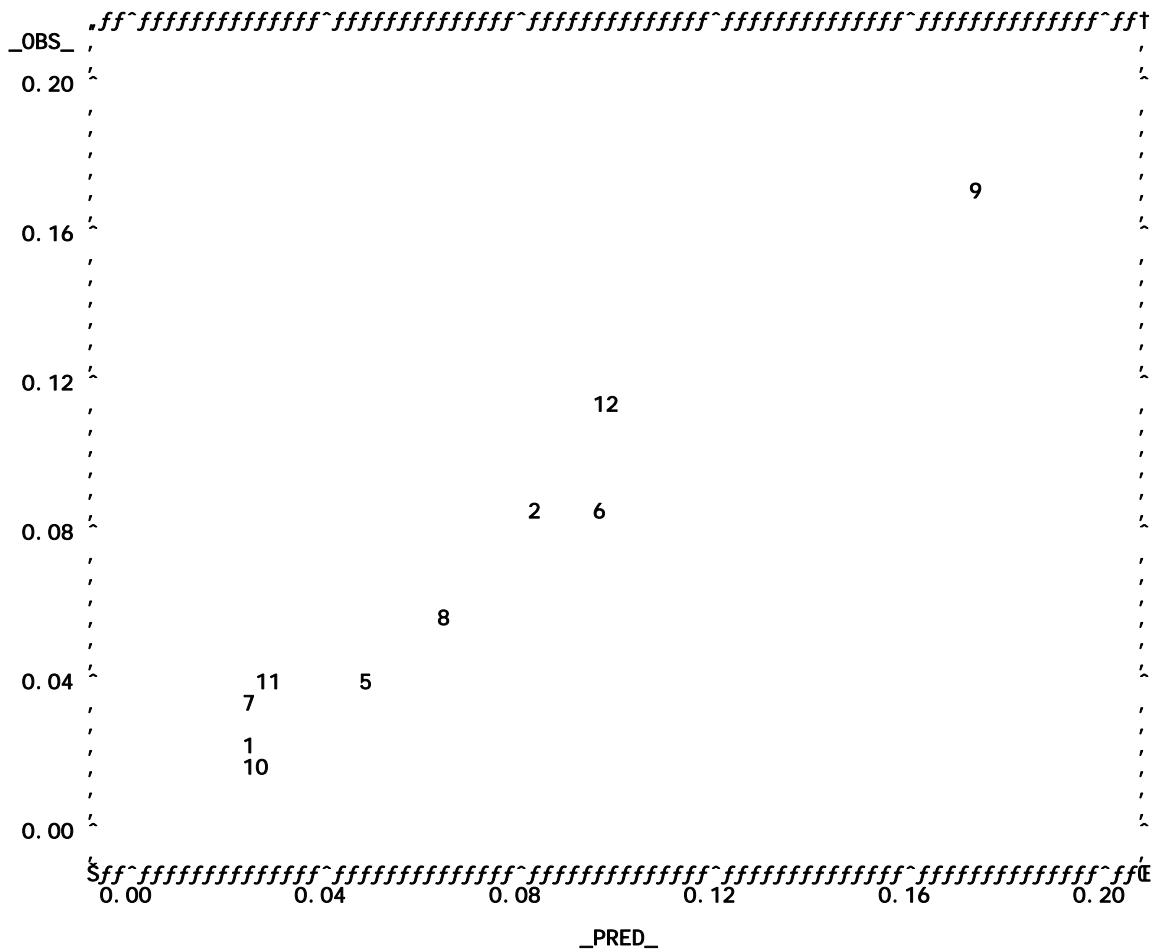
Effect	Parameter	Estimate	Standard Error	Chi-Square	Prob
INTERCEPT	1	0.0715	0.00340	441.18	0.0000
SPEED	2	-0.0480	0.00381	158.33	0.0000
	3	-0.0157	0.00432	13.17	0.0003
STEERING(SPEED=30-49)	4	0.0106	0.00409	6.73	0.0095
SEATBELT(SPEED=30-49)	5	0.0174	0.00455	14.69	0.0001
SEATBELT(SPEED=50+)	6	0.0382	0.00862	19.61	0.0000

PREDICTED VALUES FOR RESPONSE FUNCTIONS

Sample	Function Number	-----Observed-----		-----Predicted-----		Residual
		Function	Standard Error	Function	Standard Error	
1	1	0.02137931	0.00379857	0.02352052	0.00297772	-0.0021412
2	1	0.08701595	0.00601608	0.08387468	0.00568495	0.00314126
3	1	0.17335474	0.01072429	0.17335474	0.00875634	0
4	1	0.02298851	0.00927652	0.02352052	0.00297772	-0.000532
5	1	0.03966006	0.01038728	0.04902447	0.008571	-0.0093644
6	1	0.08675799	0.01902066	0.09698596	0.01485757	-0.010228
7	1	0.03264095	0.00684456	0.02352052	0.00297772	0.00912043
8	1	0.05786802	0.00743973	0.06267189	0.00680353	-0.0048039
9	1	0.17335474	0.01516643	0.17335474	0.00875634	0
10	1	0.01970443	0.00975467	0.02352052	0.00297772	-0.0038161
11	1	0.03985507	0.01177486	0.02782168	0.00904371	0.0120334
12	1	0.11299435	0.02379605	0.09698596	0.01485757	0.0160084

OBS	STEERING	SEATBELT	SPEED	_SAMPLE_	_TYPE_	_NUMBER_	_OBS_	_SEOBS_	_PRED_	_SEPRD_	_RESID_
1	absent	no	0-29	1	FUNCTION	1	0.02138	0.003799	0.02352	0.002978	-0.002141
2	absent	no	30-49	2	FUNCTION	1	0.08702	0.006016	0.08387	0.005685	0.003141
3	absent	no	50+	3	FUNCTION	1	0.17335	0.010724	0.17335	0.008756	0.000000
4	absent	yes	0-29	4	FUNCTION	1	0.02299	0.009277	0.02352	0.002978	-0.000532
5	absent	yes	30-49	5	FUNCTION	1	0.03966	0.010387	0.04902	0.008571	-0.009364
6	absent	yes	50+	6	FUNCTION	1	0.08676	0.019021	0.09699	0.014858	-0.010228
7	present	no	0-29	7	FUNCTION	1	0.03264	0.006845	0.02352	0.002978	0.009120
8	present	no	30-49	8	FUNCTION	1	0.05787	0.007440	0.06267	0.006804	-0.004804
9	present	no	50+	9	FUNCTION	1	0.17335	0.015166	0.17335	0.008756	0.000000
10	present	yes	0-29	10	FUNCTION	1	0.01970	0.009755	0.02352	0.002978	-0.003816
11	present	yes	30-49	11	FUNCTION	1	0.03986	0.011775	0.02782	0.009044	0.012033
12	present	yes	50+	12	FUNCTION	1	0.11299	0.023796	0.09699	0.014858	0.016008

Plot of _OBS_*_PRED_\$I. Symbol used is ' '.



NOTE: 2 label characters hidden.