MODELISATION D’UNE REPONSE POLYTOMIQUE:

Régression logistique ordinale

RÉGRESSION LOGISTIQUE MULTIPLE
REPONSE Y ORDINALE

La variable **Y** prend les valeurs **1**, ... , **m**, **m+1** ordonnées.

1 Le modèle

$$\text{Prob}\left(Y \leq i / x_1 \cdots x_k\right) = \frac{e^{\alpha_i + \beta_1 x_1 + \cdots \beta_k x_k}}{1 + e^{\alpha_i + \beta_1 x_1 + \cdots \beta_k x_k}}$$

pour $i = 1 \ldots m$

avec $\alpha_1 \leq \alpha_2 \leq \ \ldots \ \leq \alpha_m$

Ce modèle ne doit être utilisé que sous certaines conditions: c'est le test suivant qui permet de prendre une décision.

2 Test du modèle à rapports des chances proportionnels

Le modèle général

$$\mathrm{Prob}\left(Y \le i / x_1 \ldots x_k\right) = \frac{e^{\alpha_i + \beta_{i1}x_1 + \ldots \beta_{ik}x_k}}{1 + e^{\alpha_i + \beta_{i1}x_1 + \ldots \beta_{ik}x_k}}$$

pour $i = 1 \ldots m$

Test

$$H_0 \quad \begin{cases} \beta_{11} = \beta_{21} = \ \ldots \ = \beta_{m1} \\ \beta_{12} = \beta_{22} = \ \ldots \ = \beta_{m2} \\ \beta_{1k} = \beta_{2k} = \ \ldots \ = \beta_{mk} \end{cases}$$

$k(m-1)$ contraintes

Statistique utilisée

$$\text{Score } U\left(\hat{\beta}_{H_0}\right)' \hat{I}\left(\hat{\beta}_{H_0}\right)^{-1} U\left(\hat{\beta}_{H_0}\right)$$

$$\rightarrow \chi^2_{[k(m-1)]} \quad \text{sous } H_0$$

Remarques

→ **Si l'on ne rejette pas l'hypothèse H_0, l'utilisation de la régression logistique ordinale est recommandée.**

En revanche, en cas de rejet il est préférable d'utiliser une modélisation par logit généralisé (régression logistique multinomiale)

→ Le modèle construit en régression logistique ordinale est un modèle à rapports des chances proportionnels

$$\frac{\mathbf{Prob}(\mathbf{Y} \leq \mathbf{i} / \mathbf{x}) \;/\; \mathbf{Prob}(\mathbf{Y} > \mathbf{i} / \mathbf{x})}{\mathbf{Prob}(\mathbf{Y} \leq \mathbf{i} / \mathbf{x}') \;/\; \mathbf{Prob}(\mathbf{Y} > \mathbf{i} / \mathbf{x}')}$$

$$= \frac{\mathbf{e}^{\alpha_i + \mathbf{x}\beta}}{\mathbf{e}^{\alpha_i + \mathbf{x}'\beta}} = \mathbf{e}^{(\mathbf{x}-\mathbf{x}')\beta}$$ est indépendant de $\mathbf{i}$

EXEMPLE: ETUDE DE LA QUALITE DES VINS DE BORDEAUX

ANNEE	TEMP	SOLEIL	CHAL	PLUIES	QUALITE
1924	3064	1201	10	361	2
1925	3000	1053	11	338	3
1926	3155	1133	19	393	2
1927	3085	970	4	467	3
1928	3245	1258	36	294	1
1929	3267	1386	35	225	1
1930	3080	966	13	417	3
1931	2974	1189	12	488	3
....					
....					
1955	3247	1277	19	375	1
1956	3083	1195	5	441	3
1957	3043	1208	14	371	3

Procédure LOGISTIC : Modèle à pentes parallèles.Test de pentes

Response Profile

Ordered Value	QUALITE	Count
1	1	11
2	2	11
3	3	12

Score Test for the Proportional Odds Assumption

Chi-Square = 2.9159 with 4 DF (p=0.5720)

Procédure LOGISTIC : Modèle à pentes parallèles. Estimation des paramètres

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Standardized Estimate	Odds Ratio
INTERCP1	1	-85.5074	34.9209	5.9957	0.0143	.	.
INTERCP2	1	-80.5495	33.9650	5.6242	0.0177	.	.
TEMP	1	0.0243	0.0128	3.6125	0.0573	1.889278	1.025
SOLEIL	1	0.0138	0.00850	2.6335	0.1046	0.962717	1.014
CHAL	1	-0.0888	0.1193	0.5536	0.4568	-0.490175	0.915
PLUIES	1	-0.0259	0.0124	4.3931	0.0361	-1.304902	0.974

Procédure LOGISTIC : Modèle à pentes parallèles. Validation du modèle

Model Fitting Information and Testing Global Null Hypothesis BETA=0

Criterion	Intercept Only	Intercept and Covariates	Chi-Square for Covariates
AIC	78.647	38.158	.
SC	81.700	47.316	.
-2 LOG L	74.647	26.158	48.489 with 4 DF (p=0.0001)
Score	.	.	26.044 with 4 DF (p=0.0001)

Procédure LOGISTIC : Modèle à pentes parallèles

```
•   The LOGISTIC Procedure
•
•          Association of Predicted Probabilities and Observed Responses
•
•                    Concordant = 96.1%          Somers' D = 0.922
•                    Discordant =  3.9%          Gamma     = 0.922
•                    Tied       =  0.0%          Tau-a     = 0.633
•                    (385 pairs)                 c         = 0.961
```

Procédure LOGISTIC : Modèle à pentes parallèles: prévision des probabilités

		QUALITE	PRED	L	U
INDIVIDU 1	GR 1	2	0.00806	0.00044	0.13070
	GR 2	2	0.53626	0.19576	0.84600
INDIVIDU 2	GR 1	3	0.00037	0.00000	0.04308
	GR 2	3	0.05010	0.00335	0.45254
INDIVIDU 3	GR 1	2	0.00566	0.00022	0.13078
	GR 2	2	0.44739	0.08077	0.88179
INDIVIDU 4	GR 1	3	0.00006	0.00000	0.03586
	GR 2	3	0.00864	0.00006	0.53963
INDIVIDU 5	GR 1	1	0.44863	0.04502	0.93352
	GR 2	1	0.99144	0.72510	0.99980

Régression logistique ordinale

Toutes les options classiques de la procédure logistic telles que:

- Le choix de la fonction de lien
- la sélection de variables

sont disponibles.