LA RÉGRESSION DE POISSON

Pierre-louis GONZALEZ

La régression de Poisson

La régression de Poisson permet de modéliser des **comptages** distribués selon une loi de Poisson en fonction de variables explicatives quantitatives ou qualitatives.

Y = comptage

X_1 ... X_k Variables explicatives

Les données

Âge	Région	Y=mélanome	Population
< 35	N	61	2 880 262
35 - 44	N	76	564 535
45 - 54	N	98	592 983
55 - 64	N	104	450 740
65 - 74	N	63	270 908
> 74	N	80	161 850
< 35	S	64	1 074 246
35 - 44	S	75	220 407
45 - 54	S	68	198 119
55 - 64	S	63	134 084
65 - 74	S	45	70 708
> 74	S	27	34 233

Le modèle (paramétrisation Genmod)

$$\mu = N \exp\Bigg[\beta_0 + \begin{array}{c} <35 \\ 35-44 \\ 45-54 \\ 55-64 \\ 65-74 \\ >74 \end{array}\begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \\ 0 \end{bmatrix} + \begin{array}{c} N \\ S \end{array}\begin{bmatrix} \beta_6 \\ 0 \end{bmatrix}$$

Effectif population soumise au risque

Âge

Région

$$+ \begin{array}{c|c|c|} & N & S \\ <35 & \beta_7 & -\beta_7 \\ 35-44 & \beta_8 & -\beta_8 \\ 45-54 & \beta_9 & -\beta_9 \\ 55-64 & \beta_{10} & -\beta_{10} \\ 65-74 & \beta_{11} & -\beta_{11} \\ >74 & 0 & 0 \end{array}$$

Âge * Région

$$P(Y = y) = \frac{e^{-\mu}\,\mu^{y}}{y!}$$

Le modèle (paramétrisation Catmod)

$$\mu = N \exp\left[\beta_0 + \begin{array}{c} <35 \\ 35-44 \\ 45-54 \\ 55-64 \\ 65-74 \\ >74 \end{array} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \\ -\beta_1 \ldots -\beta_5 \end{bmatrix} + \begin{array}{c} N \\ S \end{array} \begin{bmatrix} \beta_6 \\ -\beta_6 \end{bmatrix}\right.$$

Effectif population soumise au risque

Âge

Région

$$+ \begin{array}{c|c|c|} & N & S \\ <35 & \beta_7 & -\beta_7 \\ 35-44 & \beta_8 & -\beta_8 \\ 45-54 & \beta_9 & -\beta_9 \\ 55-64 & \beta_{10} & -\beta_{10} \\ 65-74 & \beta_{11} & -\beta_{11} \\ >74 & -\beta_7 \ldots -\beta_{11} & \beta_7 + \ldots + \beta_{11} \end{array}$$

Âge * Région

$$P(Y = y) = \frac{e^{-\mu}\,\mu^{y}}{y!}$$

Remarque:

Avec la procédure GENMOD, utilisée dans l'exemple Mélanomes, le dernier paramètre pour chaque effet est fixé à 0 par défaut.

Avec la procédure CATMOD, la somme des paramètres pour chaque effet est fixée à 0.

▪ Vraisemblance

$$\varphi\left(y_1,\ldots y_s \,/\, \beta_0,\beta_1\ldots\beta_{11}\right) = \prod_{i=1}^{12} \frac{e^{-\mu_i}\mu_i^{y_i}}{y_i\,!}$$

où **i** est l’indice de la **i**ème population.

▪ Estimation

On estime les β_j en maximisant la vraisemblance.

- **Test**

$$\mathbf{H_0 : L\beta = 0}$$
$$\mathbf{H_1 : L\beta \neq 0}$$

Statistique de Wald

$$\mathbf{Q = \left(L\hat{\beta}\right)' \left[\hat{V}ar\left(L\hat{\beta}\right)\right]^{-1} L\hat{\beta}}$$

Conclusion

On rejette $\mathbf{H_0}$ si : $$Q \geq \chi^{2;\,\mathrm{rang}(L)}_{0,95}$$