

CHAPITRE 6 : LES MODELES DE COMPTAGE

SECTION 1 : Présentation du modèle

La variable endogène peut se présenter sous la forme standard suivante :

$$y = \begin{cases} 1 \\ 2 \\ 3 \end{cases}$$

On ne voit aucun changement par rapport aux chapitres précédents si ce n'est que les chiffres (1-2-3..) représentent le nombre d'occurrence d'un événement.

Les événements peuvent être :

- Le nombre de téléviseurs dans une maison,
- Le nombre d'enfants d'un couple,
- Le nombre d'élèves par classe,
- Le nombre de voyages en avion sur une année pour une personne,
- Le nombre de buts marqués dans un championnat,
- Le nombre de médailles par nation,
- Le nombre de voitures d'un ménage,
-

Quelle est la meilleure loi de probabilité pour prendre en compte le nombre d'événements dans un laps de temps fixé ?

C'est la loi de Poisson. Cette loi a été introduite en 1838 par [Siméon Denis Poisson](#) (1781–1840), dans son ouvrage « *Recherches sur la probabilité des jugements en matière criminelle et en matière civile* ».

Sa fonction de probabilité est donnée par l'expression suivante :

$$f(y) = P(Y = y) = \frac{\lambda^y e^{-\lambda}}{y!}$$

Avec Y une variable aléatoire de poisson

y un entier naturel : 0, 1, 2, 3...

λ le nombre moyen d'occurrences dans un intervalle, nombre réel strictement positif.

NB : cette fonction de probabilité n'ayant qu'un seul paramètre (λ) cela implique que $E(Y) = \text{Var}(Y) = \lambda$.

Dans notre étude cela implique que nous allons devoir estimer le modèle suivant (appelé **modèle de régression de Poisson** qui utilise la fonction génératrice des moments de la loi de poisson) de la forme :

$$E(Y) = \lambda = \exp(cste + \beta X)$$

On estime **CSTE** et β par la méthode du maximum de vraisemblance, on en déduit $\hat{\lambda}$:

$$\widehat{E(Y)} = \hat{\lambda} = \exp(\widehat{cste} + \hat{\beta}X)$$

On en déduit aussi la probabilité du nombre d'occurrences :

$$P(\widehat{Y} = y) = \frac{\hat{\lambda}^y e^{-\hat{\lambda}}}{y!}$$

SECTION 3 : Application : le nombre de médailles gagnées aux jeux olympiques

Base de données : MEDAILLES.XLS

N= ± 1610 observations (pays)

Variable endogène :

$$\text{totalmedaille} = \begin{cases} 0 \\ \text{à} \\ NN \end{cases}$$

Variables exogènes :

PAYSCODE = 1 Afghanistan, 2 Albanie.....

ANNEES = de 1960 à 1992

PAYS = nom du pays

PIB = en \$

POP = population du pays en millions

OR = nombre de médailles d'or

ARGENT = nombre de médailles d'argent

BRONZE = nombre de médailles en bronze

PAYSHOTE = 1 si le pays accueille les jeux olympiques

Le modèle est le suivant :

$$\widehat{totalmedaille} = \exp(\widehat{cste} + \widehat{\beta}_1 POP + \widehat{\beta}_2 PIB)$$

NB : POP et PIB sont pris en ln.

Avec SAS 9.4 nous obtenons :

poisson regression

The COUNTREG Procedure

Model Fit Summary	
Dependent Variable	TOTALMEDAILLE
Number of Observations	151
Missing Values	4
Data Set	WORK.NEW88
Model	Poisson
Log Likelihood	-722.33650
Maximum Absolute Gradient	3.04883E-9
Number of Iterations	6
Optimization Method	Newton-Raphson
AIC	1451
SBC	1460

Algorithm converged.

Résultats estimés des paramètres					
Paramètre	DDL	Valeur estimée	Erreur type	Valeur du test t	Approx. de Pr > t
Intercept	1	-15.887459	0.511805	-31.04	<.0001
Inpop	1	0.180038	0.032280	5.58	<.0001
Inpib	1	0.576603	0.024722	23.32	<.0001

Notre but est de trouver un modèle adéquat pour prévoir le nombre de médailles d'un pays une année donnée (ici 1988) sur la base du nombre de sa population et du niveau de son PIB par exemple :

POP 1988 RU	PIB 1988 RU	NB moyen de médailles estimées
5.72 ^{E07}	1.01 ^{E12}	26.21

En réalité le Royaume Uni a gagné 24 médailles en 1988. Le modèle ainsi spécifié est excellent.