

MATH 60604

Modélisation statistique

§ 4d - Régression de Poisson

Léo Belzile

HEC Montréal
Département de sciences de la décision

Régression de Poisson

- La régression de Poisson spécifie une loi de Poisson pour la variable réponse Y_i avec paramètre μ_i , $Y_i \sim \text{Po}(\mu_i)$, où

$$\mu_i = E(Y_i) = \text{Var}(Y_i).$$

- On considère le logarithme naturel $\ln(x)$ comme fonction de liaison,

$$g\{E(Y_i)\} = g(\mu_i) = \ln\{E(Y_i)\} = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_p X_{ip}.$$

- De manière équivalente, la moyenne de la réponse de l'individu i est

$$E(Y_i) = \mu_i = \exp(\beta_0 + \beta_1 X_{i1} + \cdots + \beta_p X_{ip}).$$

Interprétation du coefficient dans la régression Poisson

- Soit \mathbf{x} , \mathbf{x}_+ deux ensembles de variables explicatives identiques hormis pour la k e composante, respectivement x_k et $x_k + 1$.

Si $\mathbf{X} = \mathbf{x}$, le modèle liant la moyenne de Y aux prédicteurs est

$$\mu_i(\mathbf{x}) = E(Y_i | \mathbf{X} = \mathbf{x}) = \exp \left(\beta_0 + \sum_{j=1}^p \beta_j x_j \right),$$

tandis que pour $\mathbf{X} = \mathbf{x}_+$, on a plutôt

$$\mu_i(\mathbf{x}_+) = E(Y_i | \mathbf{X} = \mathbf{x}_+) = \exp \left(\beta_0 + \sum_{j=1}^p \beta_j x_j + \beta_k \right).$$

- Le rapport des deux moyennes, $\mu_i(\mathbf{x}_+)/\mu_i(\mathbf{x})$, est $\exp(\beta_k)$.
- Quand X_k augmente d'une unité, la moyenne de Y est **multipliée** par $\exp(\beta_k)$.

Ajustement d'une régression de Poisson avec proc genmod

On considère un modèle de Poisson pour le nombre d'items achetés suite au visionnement de la publicité.

Code SAS pour ajuster une régression Poisson

```
proc genmod data=modstat.intention;  
class educ revenu;  
model nachat=sexe age revenu educ statut  
      fixation emotion / dist=poisson link=log  
      lrci type3;  
run;
```

Tests du rapport de vraisemblance pour les paramètres - effets globaux

Statistique LR pour Analyse de Type 3			
Source	DDL	Khi-2	Pr > khi-2
sexe	1	10.97	0.0009
age	1	1.58	0.2089
revenu	2	65.18	<.0001
educ	2	4.11	0.1281
statut	1	7.40	0.0065
fixation	1	76.25	<.0001
emotion	1	30.61	<.0001

Cinq variables explicatives ont un effet non-nul selon les tests du rapport de vraisemblance.

Estimés des paramètres du modèle de Poisson

Analyse des paramètres estimés du maximum de vraisemblance								
Paramètre	DDL	Estimation	Erreur	Rapport de vraisemblance	Intervalle de confiance à 95%	Khi-2 de Wald	Pr > khi-2	
			type					
Intercept	1	-1.6305	0.6618	-2.9427	-0.3466	6.07	0.0137	
sexe	1	0.5361	0.1649	0.2168	0.8640	10.57	0.0011	
age	1	-0.0228	0.0183	-0.0590	0.0127	1.56	0.2115	
revenu	1	1	1.2463	0.2461	0.7712	1.7374	<.0001	
revenu	2	1	-0.1250	0.2532	-0.6213	0.3736	0.24	0.6216
revenu	3	0	0.0000	0.0000	0.0000	0.0000	.	.
educ	1	1	0.2497	0.2226	-0.1800	0.6948	1.26	0.2620
educ	2	1	0.4040	0.2044	0.0123	0.8159	3.90	0.0482
educ	3	0	0.0000	0.0000	0.0000	0.0000	.	.
statut	1	-0.4218	0.1558	-0.7291	-0.1175	7.33	0.0068	
fixation	1	0.5501	0.0614	0.4296	0.6708	80.16	<.0001	
emotion	1	0.7887	0.1396	0.5133	1.0610	31.92	<.0001	
Echelle	0	1.0000	0.0000	1.0000	1.0000			

Le paramètre d'échelle est un parce que la variance est complètement déterminée par le modèle pour la moyenne.

Interprétation des paramètres significatifs

- Les femmes font plus d'achat, en moyenne, que les hommes. Lorsque les autres variables demeurent fixes, la moyenne du nombre d'achats des femmes est $\exp(0,5361) = 1,71$ fois celle des hommes. Ainsi, le nombre d'achats moyen des femmes augmente de 71% par rapport à celle des hommes.
- L'estimé du paramètre de fixation est $\hat{\beta}_{\text{fixation}} = 0,55$. Plus le temps de fixation augmente, plus le nombre moyen d'achat est élevé. Toutes choses étant égales par ailleurs, augmenter *fixation* de un multiplie la moyenne du nombre d'achat par $\exp(0,55) = 1,73$.
- *Ceteris paribus*, le nombre moyen d'achat dans la catégorie à faible revenu achètent en moyenne 3,47 fois plus que ceux dont le revenu est élevé, une augmentation moyenne de 247%.

Qualité de l'ajustement

- La sortie SAS inclut un tableau qui contient la valeur de la log-vraisemblance (log-vraisemblance complète) et les critères d'information.
- Pour le modèle de régression de Poisson, deux statistiques, la déviance et la statistique X^2 de Pearson, servent à mesurer la qualité de l'ajustement et à déterminer si le modèle est adéquat.

Critères d'évaluation de l'adéquation			
Critère	DDL	Valeur	Valeur/DDL
Ecart	110	203.2710	1.8479
Déviance normalisée	110	203.2710	1.8479
Khi2 de Pearson	110	216.2705	1.9661
Pearson normalisé X2	110	216.2705	1.9661
Log-vraisemblance		3.2104	
Log-vraisemblance complète		-186.1639	
AIC (préférer les petites valeurs)		392.3279	
AICC (préférer les petites valeurs)		394.3462	
BIC (préférer les petites valeurs)		420.2028	

Les premières lignes sont dupliquées; c'est dû au fait que la variance du modèle de Poisson est complètement déterminée par la moyenne.