**Modèles linéaires pour actuaire (ACT-2003)** [kevin.laliberte-lapalme.1@ulaval.c](mailto:kevin.laliberte-lapalme.1@ulaval.c)a

Dépannage #6 : Modèle linéaire généralisé (GLM)

**#1**

Vous ajustez un modèle de régression de Poisson à un jeu de données sur le nombre d’éclairs frappant les antennes d’une certaine région. Vous utilisez



Ou xi1 et xi2 sont des variables décrivant des propriétés des antennes, et vi est le nombre d’orages ayant eu lieu durant la période de collecte des données dans la région de la ième antenne.

1. Aucun coefficient de régression ne se retrouve devant la variable ln(vi). Est-ce qu’il y a une façon raisonnable de faire entrer vi dans le modèle ? Expliquer
2. Selon le modèle utilisé pour ui, quel est l’effet d’une augmentation de 5% du nombre d’orages sur le nombre de fois qu’une antenne est frappée par des éclairs?
3. Si B1=0.24, quel est l’effet d’une hausse de 1 unité dans xi1 sur le nombre moyen de fois qu’une antenne est frappé par des éclairs?
4. Si la déviance pour ce modèle de Poisson est 123.4 et que celle pour le modèle



Est 120.1, est-ce que l’interaction entre les propriétés des antennes est significative?

1. Répondez encore une fois à la question (b) et (c) mais sous le modèle avec intéraction si B12=-0.01 et pour les valeurs xi2=1 et xi2=10

**#2**

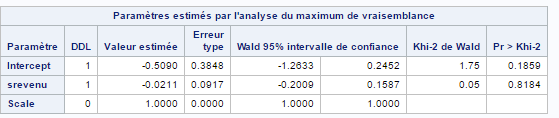
Vous devez analyser les résultats d’une étude de marché ayant généré les données suivantes :

|  |  |  |
| --- | --- | --- |
|  | Consomme le produit ABC | |
| Revenu familial | Oui | Non |
| Faible (1) | 10 | 25 |
| Moyen (3) | 15 | 20 |
| Élevé (5) | 20 | 15 |
| Très élevé (6) | 5 | 30 |

1. Si des raisons de marketing suggèrent les scores 1, 3, 5 et 6 pour les revenus faibles, moyens, élevés et très élevés, respectivement, est-ce qu’un modèle de régression logistique avec le score associé au revenu entrant comme variable exogène continue serait approprié ici ?
2. Ajustez le modèle de régression logistique



, où x est le score associé au revenu donné en (i) et g(·) est le lien logit, aux données de ce tableau. Vous ajustez ensuite un modèle avec un lien logit mais où revenu est catégorique (significative). Vous obtenez le tableau suivant avez SAS;



Est-ce que le modèle avec revenu numérique semble bien s’ajuster aux données ?

1. Ajuster le modèle en ii) avec le logiciel R

**#3**

2484 personnes ont répondu à un sondage dans lequel elles devaient indiquer la fréquence à laquelle elles ronflent et leurs antécédents cardiaques. Les réponses ont été regroupées selon la fréquence de ronflement :

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | Crise de cœur | | Proportion de oui |
| Ronflement | Score | Oui | Non |
| Jamais | 0 | 24 | 1355 | 0.0174 |
| Occasionnellement | 2 | 35 | 603 | 0.0549 |
| Presque chaque nuit | 4 | 21 | 192 | 0.0986 |
| Toutes les nuits | 5 | 30 | 224 | 0.118 |

On veut analyser ces données à l’aide d’un modèle de régression logistique afin de déterminer s’il existe un lien entre la fréquence de ronflement et le risque d’attaque cardiaque. Comme vous êtes une machine en programmation, vous obtenez les tableaux suivant afin de comparer les fonctions de lien identité, logit et probit :

|  |  |  |  |
| --- | --- | --- | --- |
|  | Test d’ajustement du modèle | | |
| Fonction lien | Déviance standardisée | ddl | Seuil observé |
| Identité | 0.0692 | 2 | 0.9559 |
| Logit | 2.8089 | 2 | 0.2455 |
| Probit | 1.8716 | 2 | 0.3922 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Source de ronflement | Proportion observée | Probabilité prédite par le modèle | | |
| Linéaire | Logistique | Probit |
| 0 | 0.017 | 0.017 | 0.021 | 0.020 |
| 2 | 0.055 | 0.057 | 0.044 | 0.046 |
| 4 | 0.099 | 0.096 | 0.093 | 0.095 |
| 5 | 0.118 | 0.116 | 0.132 | 0.131 |

Parmi les fonctions de lien identité, logit et probit, laquelle est la plus adéquate pour modéliser ces données ?

**Code information du dépannage 5**

SAS

data transp;

input rejet x @@;

datalines;

5 1 1 1 1 1 3 1 1 1 2 1 2 1 3 1 6 1 4 1

7 0 5 0 2 0 6 0 3 0 4 0 2 0 5 0 1 0 3 0

;

run;

proc genmod data=transp;

model rejet = x / dist=Poisson link=log;

run;

\end{SASinput}

\begin{SASoutput}[,fontsize=\scriptsize, frame=single,framesep=3mm]

The GENMOD Procedure

R

new<-c(5,1,1,3,1,2,2,3,6,4)

old<-c(7,5,2,6,3,4,2,5,1,3)

fit1<-c(rep(0,10))

fit2<-c(rep(1,10))

glm(new~fit1 ,family=poisson(link=log))

summary(glm(new~fit1,family=poisson(link=log)))

glm(old~fit2 ,family=poisson(link=log))

summary(glm(old~fit2,family=poisson(link=log)))

donnee<-data.frame( new,old,fit2,fit1)

glm(new~fit2+fit1 ,data=donnee,family=poisson(link=log))

summary(glm(new~fit2+fit1 ,data=donnee,family=poisson(link=log)))

anova(glm(new~fit2+fit1 ,data=donnee,family=poisson(link=log)))

glm(old~fit2+fit1 ,data=donnee,family=poisson(link=log))

summary(glm(old~fit2+fit1 ,data=donnee,family=poisson(link=log)))

anova(glm(old~fit2+fit1 ,data=donnee,family=poisson(link=log)))

glm(new+old~fit2+fit1 ,data=donnee,family=poisson(link=log))

summary(glm(new+old~fit2+fit1 ,data=donnee,family=poisson(link=log)))

anova(glm(new+old~fit2+fit1 ,data=donnee,family=poisson(link=log)))