

MODELES LOG-LINEAIRES SOUS

La variable dépendante n'est plus une variable binaire, mais une variable de comptage. Pour effectuer une régression de Poisson, on utilise toujours la commande glm, avec l'argument family=poisson.

OBJECTIF

Nous allons analyser le jeu de données sur le comptage de crises épileptiques traité initialement par Thall et Vail (1990) et contenu dans le package « MASS ».

La variable à expliquer est le nombre de crises épileptiques compté sur une période de deux semaines et ceci quatre fois (quatre périodes de deux semaines = période 1, période 2, période 3 et période 4) pour chacun des 59 patients de l'étude.

Les patients sont divisés en deux groupes : un groupe de patient recevant le traitement Progabide et un groupe de patient recevant un Placebo. Le nombre de crises épileptiques sur une période de 8 semaines avant la prise du médicament/placebo (période de référence) est également connu et reflète le comptage de référence.

Notre étude visera à expliquer l'effet du médicament Progabide (par rapport à un Placebo) contre la survenue de crises partielles d'épilepsie.

CHARGEMENT PACKAGE « MASS »

Chargement du package (à chaque utilisation) :

```
library(MASS)
```

PRESENTATION / LECTURE DES DONNEES

Les données à analyser (**epil**) contiennent les informations sur une cohorte de 59 patientes.

```
head(epil)
tail(epil)
names(epil)
```

Les variables sont définies ci-dessous :

y	nombre de crises épileptiques (comptage) sur une période de 2 semaines
trt	traitement : placebo ou progabide
base	nombre de crises épileptiques sur une période de 8 semaines précédant l'inclusion (= pendant la période de référence)
age	âge du sujet (en année)

v4	variable indicatrice de la période 4 : 0 ou 1
subject	numéro du sujet de 1 à 59
period	période 1 à période 4
lbase	log (nombre de crises épileptiques) pour la période « baseline » (centré pour avoir une moyenne 0)
lage	log (âge) (centré pour avoir une moyenne 0)

OBJECTIF DE MODELISATION

On souhaite modéliser le nombre de crises épileptiques (y) compte tenu des covariables démographiques (age), cliniques ($base$) et du traitement administré (trt) sur les 59 patients de la cohorte étudiée.

ANALYSE EXPLORATOIRE

```
str(epil)
summary(epil)
data=epil
```

Les données dichotomiques ne sont pas traitées comme des variables qualitatives. On les transforme donc en type facteur :

```
data_col = c(5,6,7)
for (i in 1:3){
  data[,data_col[i]] = factor(data[,data_col[i]])
}

str(data)
summary(data)
```

Description des données qualitatives (exemple) :

```
table(data[,2])
table(data[,2],data[,7])
```

Description des données quantitatives (exemple) :

```
V4=data[,5]
data2=data[V4==1,]
```

Variable **base** = nombre de crises épileptiques pdt la période de référence

```
median(data2[,3])
min(data2[,3])
max(data2[,3])
```

Variable **age** = age

```
median(data2[,4])
min(data2[,4])
max(data2[,4])
```

Discrétisation de la variable **base** :

```
borne = c(0,22,max(epil[,3]))
base_cl = cut(epil[,3], breaks = borne, include.lowest = TRUE)
```

```
table(base_cl)
```

Discrétisation de la variable **age** :

```
borne = c(0,28,max(epil[,4]))  
age_cl = cut(epil[,4], breaks = borne, include.lowest = TRUE)  
table(age_cl)
```

Représentation graphique / exemple 1 :

```
y=epil[,1]  
trt=epil[,2]  
base=epil[,3]  
age=epil[,4]  
V4=epil[,5]  
subject=epil[,6]  
periode=epil[,7]  
plot(periode,y,col = "black", xlab = "",ylab = "")
```

Représentation graphique / exemple 2 :

```
y1=epil[trt=="progabide",1]  
periode1=epil[trt=="progabide",7]  
y2=epil[trt=="placebo",1]  
periode2=epil[trt=="placebo",7]  
  
plot(periode1,y1,col = "red", xlab = "",ylab = "")  
points(periode2,y2,col = "blue", xlab = "",ylab = "")
```

MODELE LOG-LINEAIRE

Modèle log-linéaire sans interaction :

```
modele1 = glm(y~periode+age_cl+trt+base,family=poisson,data=epil)  
summary(modele1)  
exp(cbind(RR1 = coef(modele1), confint(modele1)))  
AIC(modele1)  
BIC(modele1)
```

Modèle log-linéaire avec interaction :

```
modele2 = glm(y~periode+age_cl+trt+base+base:trt,family=poisson,data=epil)  
summary(modele2)  
AIC(modele2)  
BIC(modele2)
```

Test de l'ajout d'une interaction dans le modèle :

```
lrtest(modele1,modele2)
```

EXERCICE 1

1/ Réfléchir à l'interprétation des coefficients (en terme de ratio rate = rapport de taux)

2/ Réfléchir à l'interprétation de la modélisation

MODELE LOG-LINEAIRE POUR TABLES DE CONTINGENCE

Table de contingence pour les variables :

- traitement
- nombre de crises épileptiques (en classe) pdt la période de référence
- âge (en classe)

```
mytable=xtabs(~trt+base_cl+age_cl)
ftable(mytable)
freqdata=data.frame(mytable)
freqdata
```

Modèle log-linéaire pour tables de contingences :

```
modele=glm(Freq~trt*base_cl*age_cl,family=poisson,data=freqdata)
modele
```

Analyse des interactions entre les variables :

```
anova(modele, test="Chisq")
```

EXERCICE 2

1/ Quelles sont les interactions significatives ?

2/ Relancer les commandes précédentes avec seulement les variables traitement et âge (en 3 classes : <25, [25, 30[et ≥30) et comparer les résultats obtenus entre le test du χ^2 d'indépendance et le test du paramètre d'interaction dans le modèle log-linéaire.