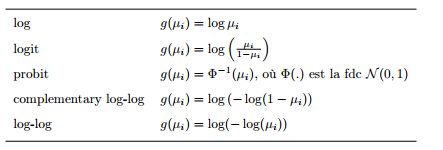
**Modèles linéaires pour actuaire (ACT-2003)** [kevin.laliberte-lapalme.1@ulaval.c](mailto:kevin.laliberte-lapalme.1@ulaval.c)a

Dépannage #7 : Modèle linéaire généralisé (GLM)

Rappel fonction lien



**#1**

Nous allons étudier un portefeuille d’assurance automobile. Chaque cellule contient les données agrégées de conducteurs ayant des valeurs identiques pour les facteurs de risques suivant :

* *genre* : 1 = femme, 2 = homme
* *région* : 1 = rurale, 2 = autre, 3 = urbaine
* *type de la voiture* : 1 = petite, 2 = moyenne, 3 = grande
* *emploi* : 1 = fonctionnaire/actuaire..., 2 = entre-deux, 3 = ‘dynamique’

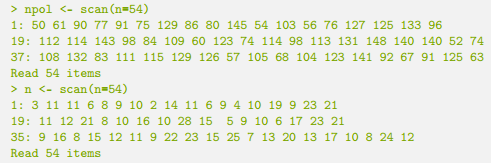
Pour chaque cellule, nous connaissons :

* *npol* : nombre total d’assurés dans cette cellule
* *n* : nombre total de sinistres observés durant l’année précédente

Les observations sont ordonnées de telle sorte que

* la variable genre est égal à 1 pour les 27 premières observations, et 2 pour les 27 dernières observations
* la variable région a trois modalités qui se produisent par blocs de 9 observations
* la variable type a trois modalités qui se produisent par blocs de 3 observations
* la variable emploi a trois modalités qui se répètent jusqu’à atteindre la 54 ème observation.

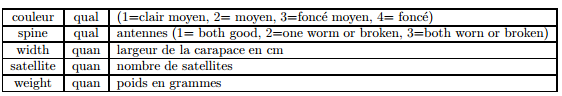
Les modalités des variables catégorielles peuvent être reconstruits en utilisant la fonction gl(). La commande gl(n,k,N) produit un vecteur de longueur N contenant les catégories numérotées 1 à n par blocs de longueur k. La commande gl(n,k,N) est équivalente à as.factor(rep (1: n, each = k, length = N)). Le nombre moyen de sinistres par contrat n/npol dans chaque cellule est la quantité d’intérêt.

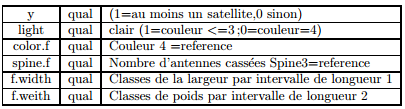


Modéliser cette régression par un glm de poisson avec un lien log.

**#2**

On étudie une espèce de limules vivant sur la côte est des États-Unis nommé crabe en fer à cheval. Le fichier http://www.math.unicaen.fr/~kauffman/data/crab.txt contient cinq colonnes et 173 lignes. Les variables sont



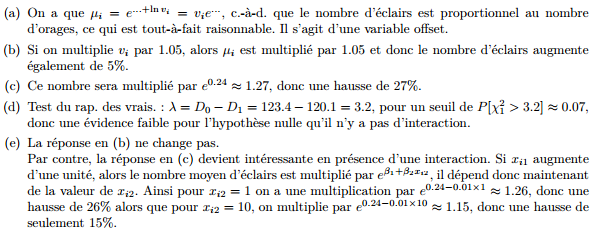


1. Est-ce qu’une régression linéaire multiple serait appropriée pour modéliser cette population? Expliquer
2. Est-ce qu’un GLM de poisson avec lien identité, probit ou log serait approprié ? Expliquer

#3

Je réponds aux questions du tp

**Réponse #1 depannage.6**



**Code dépannage 6**

**R**

score<-c(1,3,5,6)

yes<-c(10,15,20,5)

no<-c(25,20,15,30)

tot<-yes+no

probb<-yes/tot

summary(glm(probb~score,family = binomial(link="logit")))

anova(glm(probb~score,family = binomial(link="logit")))

predict(glm(probb~score,family = binomial(link="logit")))

**SAS**

data model;

input srevenu noui n;

datalines;

1 10 35

3 15 35

5 20 35

6 5 35

;

run;

proc genmod data=model;

model noui / n = srevenu / dist=bin link=logit type3;

output out=num pred=pnum;

run;

proc genmod data=model;

class srevenu;

model noui / n = srevenu / dist=bin link=logit type3;

lsmeans srevenu / ilink;

output out=catego pred=pcat;

run;