

## Examen Modèles Linéaires Généralisés

Seconde session - septembre 2013 - durée : 2h - 2 exercices

Documents autorisés : cours et tables statistiques

*Note : Tout calcul doit être expliqué et découler d'une formule donnée avec précision.*

### Exercice 1

On considère un modèle log-linéaire pour expliquer les effectifs observés dans chacune des cellules obtenues par croisement de trois facteurs  $A$ ,  $B$  et  $C$  à  $J$ ,  $K$  et  $L$  niveaux respectivement.

1. Dans un premier temps, on estime le modèle saturé :

(a) Quelles variables contient-il ? Combien de paramètres sont à estimer ?

(b) Que vaut sa déviance ? Pourquoi ?

2. Parmi les modèles suivants, lesquels sont estimables ? Pourquoi ?

A B A\*B

A B B\*C

A B C A\*B

A|B|C

A\*B\*C

A B A\*B A\*C B\*C

3. Après une sélection descendante, on trouve un modèle composé des trois facteurs  $A$ ,  $B$  et  $C$  et de l'interaction  $A * C$ , avec les p-values respectives : 0.001, 0.005, 0.08 et 0.002.

(a) Doit-on retirer le facteur  $C$  de ce modèle ? Pourquoi ?

(b) Que signifie concrètement l'interaction significative entre  $A$  et  $C$  ?

### Exercice 2

On considère la loi Normale  $N(\mu, \sigma^2)$ , où  $\mu$  et  $\sigma^2$  sont deux paramètres inconnus à estimer. On réalise un échantillon de taille  $n$ , noté  $(y_1, \dots, y_n)$ .

On rappelle que la densité d'une loi  $N(\mu, \sigma^2)$  est définie par :

$$f(y; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2}(y - \mu)^2\right]$$

1. Ecrire la log-vraisemblance du modèle considéré.

2. Déterminer le score du modèle, c'est-à-dire l'expression de la dérivée première de la log-vraisemblance par rapport à  $\mu$ .

3. En déduire  $\hat{\mu}$ , estimateur du maximum de vraisemblance de  $\mu$ .

4. Vérifier que  $\hat{\mu}$  est un estimateur sans biais de  $\mu$ .

5. Dans le cadre du modèle linéaire généralisé, on suppose que les  $y_i$  ( $i = 1, \dots, n$ ) sont distribuées selon une loi  $N(\mu_i, \sigma^2)$ .

(a) Montrer que cette loi appartient à la famille des lois exponentielles, avec :

$$\theta_i = \mu_i$$

$$b(\theta_i) = \mu_i^2/2$$

$$a_i(\phi) = \phi = \sigma^2$$

(b) Vérifier les deux égalités suivantes :

$$b'(\theta_i) = E(Y_i) = \mu_i \qquad \text{Var}(Y_i) = a_i(\phi)b''(\theta_i)$$

### Exercice 3

Une étude a été réalisée sur 42 patients hypercholestérolémiques, pour évaluer l'effet de trois traitements au bout d'une année. Parmi ces 42 patients, 15 d'entre eux sont guéris au bout d'un an (leur taux de cholestérol est revenu dans les normes). Pour chaque patient, on dispose également de son poids (en kg) et du traitement suivi (A, B et C). On souhaite étudier l'effet éventuel du poids et du traitement sur la probabilité de guérison du patient, en mettant en œuvre un modèle de régression logistique.

1. En l'absence d'effet de variables explicatives, que vaut la probabilité de guérison ?
2. Une première régression logistique en fonction du poids est mise en œuvre (notée M1); quelques résultats vous sont fournis en annexe 3.1.
  - (a) Ecrivez le modèle M1 en précisant les notations utilisées.
  - (b) Testez l'effet du poids sur la probabilité de guérison.
3. Une autre régression logistique a été estimée en fonction du poids et du traitement (notée M2); quelques résultats vous sont fournis en annexe 3.2.
  - (a) Ecrivez le modèle M2 associé aux résultats fournis.
  - (b) Comment expliquez-vous qu'il n'y ait pas de paramètre associé au traitement A ?
  - (c) Calculez la déviance relative de ce modèle. Commentez.
  - (d) Comparez les deux modèles M1 et M2 en utilisant un test d'hypothèse. Quel modèle sélectionnez-vous ? Que pouvez-vous en déduire sur l'effet du traitement ?
4. A partir du modèle sélectionné, interprétez les effets du (ou des) variable(s) explicative(s).

#### Annexe 3.1 : Régression logistique en fonction du poids

	AIC		54.033
	-2 Log L		50.033
		Valeur	Erreur
Paramètre	DDL	estimée	type
Intercept	1	-2.6141	1.0722
Poids	1	0.0526	0.0259

#### Annexe 3.2 : Régression logistique en fonction du poids et du traitement

	AIC		47.012
	-2 Log L		39.012
		Valeur	
Paramètre	DDL	estimée	
Intercept	1	-4.3896	
Poids	1	0.0655	
Traitement B	1	0.7377	
Traitement C	1	3.0949	