Conservatoire national des arts et métiers : STA 112
Examen
27 juin 2013. 18h00-20h00

Documents (papier ou électronique) et machines à calculer autorisés

1 Questions autonomes

Justifier et détailler vos réponses

1. On considère 2 séquences S_1 : CCMMMD et S_2 : CMMDMD qui retracent le statut marital par 5 années :(Celibat, Marié, Divorcé) . On suppose que les coûts d'une insertion d'une substitution et d'une délétion sont identiques. Proposer 2 transformations de S_1 en S_2 que vous classerez par coût croissant.

2. Quel est d'après vous l'avantage d'utiliser des approches fondées sur l'analyse des données chez les cas seulement?

3. Soit C(h) la covariance d'un processus stationnaire d'ordre 2 à une distance h : Montrer que $C(h) \leq C(0)$.

4. Soit $Z(X)$ une variable spatialité . Est-il possible qu'en utilisant un krigeage simple, les poids estimés soient tous nuls ?

5. On considère les 6 prélèvements de plomb espacés de 1 mètre sur une ligne. Calculer le variogramme empirique à 0, 1, 2 mètres.

4	7	8	10	15	20

6. Soit Y_i le nombre observé de cas dans l'unité géographique i, E_i le nombre attendu de cas. On suppose que $Y_i \sim Poisson(\theta_i E_i)$.
   (a) Calculer la variance de $\hat{\theta}_i$ (estimateur du maximum de vraisemblance).
   (b) Sur les données de cas de cancers oral en écosse (56 unités géographiques), on obtient les résultats suivants la valeur moyenne du SMR vaut 1.57, la variance des SMR vaut 1.71. Le modèle proposé pour les Y_i est-il compatible avec ces estimations ?
   (c) Quel modèle proposeriez vous pour prendre en compte la nature spatiale des données ?

2 Exercice

La réponse anticorps de l'enfant se mesure en concentration d'anticorps (mg/l). On souhaite expliquer le comportement immunitaire de 50 nouveaux nés de 0 à 18 mois. On note $Y_{i,t}$ la concentration d'anticorps de l'individu i au temps t (t variant de 0 à 18 mois).

Le tracé des log-concentrations de l'échantillon d'enfants est représenté en Figure (1).

1. On note le modèle de régression $M_0 : Y_{i,t} = \alpha + \beta t + \varepsilon_{i,t}$. Ce modèle est il adapté pour décrire les données ?

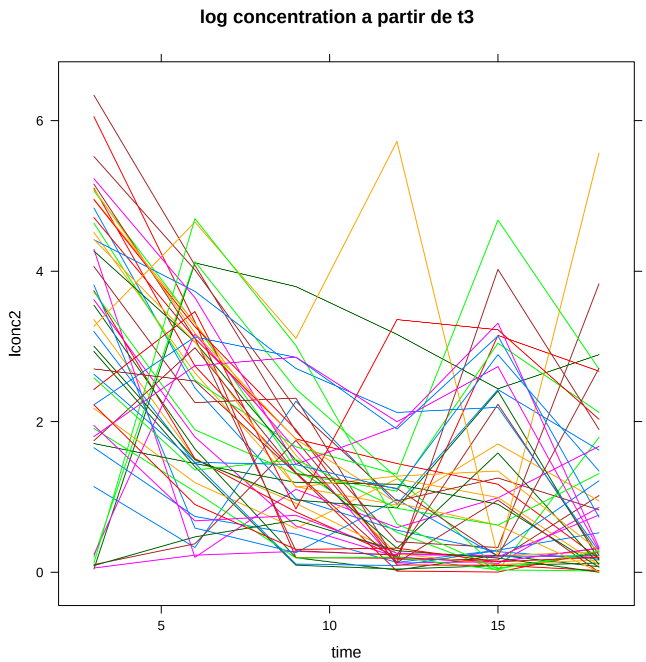

Figure 1: Log concentration en fonction du temps

2. On note le modèle de régression M_1 : $Y_{i,t} = \alpha + \alpha_i + \beta t + \varepsilon_{i,t}$. Ce modèle est il adapté pour décrire les données ?

3. On note le modèle de régression M_2 : $Y_{i,t} = \alpha + \alpha_i + (\beta + \beta_i)t + \varepsilon_{i,t}$. Ce modèle est il adapté pour décrire les données ?

On notera $\varepsilon_{i,t} \sim N(0, \sigma^2)$ le terme d'erreur aléatoire, $\alpha_i \sim N(0, \mu^2)$ et $\beta_i \sim N(0, \nu^2)$

L'estimation des paramètres du modèle 2 nous fournit :

$\hat{\alpha}$	$\hat{\mu}$	$\hat{\beta}$	$\hat{\nu}$	$\hat{\sigma}$
3.14	1.18	-0.14	0.01	1.22

Enfant	(Intercept)	time
1	0.62	-0.06
2	-0.12	-0.01

Table 1: Estimation du modèle 2 et des effets aléatoires

1. Calculer un intervalle de prédiction à 95% de la log-concentration pour les enfants 1 et 2.

2. Calculer un intervalle de prédiction à 95% de la log-concentration pour l'échantillon de 50 enfants.