

**Devoir 2**  
STT6415 - Régression  
Échéance: mardi 16 février.

Dans les problèmes énoncés ici, les données correspondent aux données du diabète sucré insulino-dépendant. Les problèmes concernent la régression (non paramétrique) de  $\log(\text{C-peptide})$  en fonction de l'âge et du déficit de base.

1. Obtenez la régression de  $\log(\text{C-peptide})$  en fonction d'âge avec une fonction de lissage par noyaux. Choisissez la valeur optimale du paramètre de lissage en utilisant le  $C_p$  de Mallou.
  - (a) Combien de degrés de liberté ont été utilisés par le lissage?
  - (b) Obtenez les bandes de confiance 95%. Calculez les bandes analytiques et les bandes basées sur le bootstrap.
  - (c) Faites un ajustement par un polynôme cubique. Soit  $f_1$  cette ajustement. Faites le test d'hypothèses que  $f_1$  est biaisée. Quelle est votre conclusion au niveau de signification  $\alpha = 0.05$  ?  
**Allusion:** Choisissez une fonction de lissage  $f_2$  qui fait très peu de lissage, de telle façon que son biais soit négligeable.
2. Considérez maintenant la régression de  $\log(\text{C-peptide})$  en fonction de l'âge et du déficit de base.
  - (a) Faites l'ajustement d'une modèle additif. Obtenez les splines de régression pour les données. Utilisez cinq noeuds positionnés dans les percentiles correspondants. Choisissez le meilleur modèle en utilisant les critère du  $C_p$  de Mallou. Faites le graphique du  $C_p$  de Mallou.
  - (b) Faites l'ajustement d'une modèle additif en utilisant les splines de lissage. Choisissez le modèle en utilisant les critère de validation croisée généralisée. Quels sont les degrés de liberté associés au meilleur modèle? Faites un graphique des courbes avec les bandes de confiance de niveau 95%.
  - (c) Quelle fonction de lissage semble effectuer une meilleur ajustement? Pour le modèle choisi: faites l'analyse de variance associée au modèle.