

TP: A rendre au plus tard le 29 mai 2009 (par binôme)

Sur le site de l'Université *Carnegie Mellon*, à l'adresse internet:

http://lib.stat.cmu.edu/datasets/Plasma_Retinol

vous trouvez les données sur une étude des concentrations du rétinol et de la bêta-carotène dans le plasma.

- 1) Faire une analyse descriptive des données.
- 2) Les variables "BETAPLASMA" et "RETPLASMA" suivent-elles une loi normale?

Modélisation:

I. On veut d'abord modéliser la variable "BETAPLASMA".

I.1. Trouver le modèle de régression optimal pour "BETAPLASMA" fonction des variables: *Age, Quetelet, Fat, Fiber, Alcohol, Cholesterol, Betadiet, Retdiet*. Analyser les résidus. Conclusion.

I.2. Trouver le modèle d'analyse de variance (optimal) pour "BETAPLASMA" fonction des variables: *Sex, Smokstat, Vituse*. Analyser les résidus. Conclusion.

I.3. Comparer les modèles obtenus aux questions **I.1.** et **I.2.**

II. On modélise maintenant la variable "RETPLASMA".

II.1. Trouver le modèle de régression optimal pour "RETPLASMA" fonction des variables: *Age, Quetelet, Fat, Fiber, Alcohol, Cholesterol, Betadiet, Retdiet*. Analyser les résidus. Conclusion.

II.2. Trouver le modèle d'analyse de variance (optimal) pour "RETPLASMA" fonction des variables: *Sex, Smokstat, Vituse*. Analyser les résidus. Conclusion.

II.3. Comparer les modèles obtenus aux questions **II.1.** et **II.2.**

Le code et les sorties du logiciel R sont à mettre dans l'Annexe de la rédaction.