

DM 2 TIAD - M1 Ingé. Math.

La programmation est à réaliser sous SAS et R. La table SAS est téléchargeable depuis <http://math.univ-lille1.fr/~tran/enseignements.html>

Les codes sont à envoyer à : chi.tran@univ-lille1.fr. Merci de rendre une copie et un code par binôme, avec le nom des étudiants en commentaire au début du programme.

Dans tous les tests demandés, le seuil sera 5%. Pour les tests lus dans les sorties SAS, on donnera la valeur numérique de la statistique de test et la p-valeur associée.

Exercice 1 (Rémissions pour des cancers)

On souhaite étudier les risques de rémission liés à un cancer, à partir de données de Lee (1974). Ces données sont dans la table `Remission`. Dans cette table, la variable `remiss` vaut 0 si le patient n'a pas eu de rémission et 1 sinon. Les 6 autres variables `cell smear infil li blast temp` sont des facteurs susceptibles de provoquer une rémission.

1. Faire des statistiques basiques pour décrire chacune des variables prises séparément.
2. Faire une régression logistique pour expliquer le risque de rémission en fonction des différents facteurs. On utilisera une `proc logistic` sous SAS.
3. En utilisant la commande / `selection=stepwise slentry=0.3 slstay=0.35` réaliser une sélection de variable. Expliquer ce que fait la commande et commenter le résultat de la régression.

Exercice 2 (ACP sous R ou SAS)

La table `crime.txt` nous fournit les taux de crime en 1977 (pour 100 000 individus) dans 7 catégories et pour chacun des 50 états américains. Réaliser une analyse en composante principale pour décrire les données et mieux comprendre la délinquance américaine en 1977.