**Master Pro M2, SITN, Université Claude Bernard, Lyon1**
Analyse de données
année 2006-2007

**PROJET- ANALYSE DE DONNEES**
à rendre au plus tard le 16 janvier 2007

**Problème 1.** *Mortalité infantile et durée de vie.*
A l'adresse internet

http://gsociology.icaap.org/dataupload.html

télécharger le fichier *http://gsociology.icaap.org/data/PD-Data.xls.*
**1)** Faire une étude déscriptive des données.
**2)** En ce qui concerne les données mesurées en 2003, il y a-t-il des pays qui ont le même profil?
**3)** En ce qui concerne les données mesurées en 2003, peux-t-on faire des sous-groupes de pays? Si oui, quelles variables sont importantes?
**4)** Comparer les données mesurées en 1980 et celles mesuréss en 2003.

**Problème 2.** *Pollution de l'air en $SO_2$.*
A l'adresse internet

http://lib.stat.cmu.edu/DASL/Datafiles/AirPollution.html

vous trouvez des données en ce qui concerne 41 villes des Etats-Unis. On fait 3 catégories de villes. Si la pollution moyenne d'une ville est inférieure à 50 $\mu g/m^3$ alors elle est dans le groupe 0, si elle appartient à l'intervalle [50, 70] alors elle est dans le groupe 1 et pour $> 70\mu g/m^3$ dans 2.
Quelles variables quantitatives peuvent prévoir les 3 catégories? Avec quel taux d'erreur?

**Problème 3.** *Pollution de l'air.*
A l'adresse internet

http://www-unix.oit.umass.edu/ statdata/statdata/

à la rubrique "Cluster Analysis" télécharger le fichier *pollution*.
Décrire le problème. Etudiez si la pollution de l'air dépend des conditions météo.

**Problème 4.** *Sondage habitants Montana.*
A l'adresse internet

http://lib.stat.cmu.edu/DASL/Datafiles/montanadat.html

vous trouvez les résultats d'une enquête auprès des habitants de Montana (Etats Unis). Commentez ces résultats.