

PROJET COURS DONNEES CATEGORIELLES STA212 ANNEE 2015

I. NATURE DU PROJET :

Le projet peut être réalisé sous deux formes.

1. Etude détaillée d'une procédure du logiciel SAS, ou d'un autre logiciel (R...)
2. Etude sur des données à caractère professionnel

1. ETUDE DETAILLEE D'UNE PROCEDURE LOGICIEL

Le projet consistera en une présentation de la méthode et de la procédure analysées, avec un «manuel» simple d'utilisation, et une application à des données réelles (autres que les jeux d'essais fournis avec le logiciel).

Les procédures suivantes du logiciel SAS peuvent être étudiées :

CATMOD (Aspects non étudiés en cours)

DISCRIM (Méthodes non paramétriques)

GENMOD (Options non étudiées en cours)

SURVEYLOGISTIC

PROBIT

QLIM (disponible dans SAS ETS)

Méthodes de traitement de données catégorielles avec SAS Enterprise miner

Certains thèmes, par exemple, la construction de modèles TOBIT, peuvent utiliser plusieurs procédures (QLIM et LIFEREG)

2. ETUDE SUR DES DONNEES A CARACTERE PROFESSIONNEL.

Le projet consistera à utiliser les méthodes présentées en cours dans le cadre d'une application professionnelle : exemple construction d'un score.

Il s'appuiera sur ou plusieurs des thèmes étudiés en cours :

Méthodes d'analyse discriminante

Méthodes de segmentation

Régression logistique

Régression de Poisson

Modèles log-linéaires

Modèles à réponse multinomiale

Modèles linéaires généralisés

Régression logistique PLS

II. PROCEDURE DE CHOIX

Chaque élève, régulièrement inscrit au cours, me propose par courrier électronique pierre-louis.gonzalez@cnam.fr la description (10 à 30 lignes) du sujet qu'il souhaite traiter. Soit ce sujet est validé, soit il est refusé car il s'agit, par exemple, d'une procédure logicielle déjà choisie par d'autres élèves. Dans ce cas l'élève doit faire une nouvelle proposition.

Les projets sont individuels. Les projets réalisés à deux ne sont pas acceptés.

III. PRESENTATION DU RAPPORT

La rédaction du rapport obéira aux règles classiques de présentation d'un projet: introduction présentant le problème posé, rédaction mettant en évidence la démarche choisie et les apports des analyses, conclusion. S'y ajoutera une notice d'utilisation de la procédure, pour ceux qui ont choisi la présentation d'une procédure d'un logiciel. Les parties importantes du listing (graphiques ou aides à l'interprétation) pourront être incorporées au texte ou placées en annexes. La taille du rapport doit être de 20 à 30 pages en ce qui concerne la partie rédigée; s'y ajoutent les extraits des listings nécessaires à la compréhension du texte.

On joindra au rapport, sous forme papier, une clé USB ou un CD-Rom contenant les fichiers des données, les programmes utilisés et le rapport (format Word). Ne pas oublier d'inscrire votre nom sur le CD-Rom et de le fixer à votre rapport. Les rapports transmis seulement sous forme électronique ne seront pas acceptés.

IV. REMISE DU TRAVAIL

La date de remise du projet est fixée au **28 septembre 2015 au plus tard.**

Aucun délai supplémentaire ne sera accordé.

Les projets peuvent être déposés dans la boîte aux lettres de la chaire de Statistique Appliquée (Accès 17 rez-de-chaussée) ou m'être adressés : Pierre-Louis GONZALEZ, Département IMATH Statistique Appliquée, 2D5000, CNAM, 292 rue St-Martin 75141 Paris Cedex 03.

Sur le rapport devront figurer une adresse électronique et un numéro de téléphone où l'on peut joindre l'élève.

Il n'y a pas de soutenance prévue. En revanche l'élève peut être convoqué pour apporter des précisions sur le travail qu'il a remis.

Les notes seront disponibles quatre semaines après les dates limites de remise des projets.