

## Programmation Statistique 2 - Feuille TP 2

STID 2-ième année (2018/2019)

### Partie 1

#### Présentation des données

Les données étudiées proviennent d'une Caisse d'Allocation Familiale. Les individus sont des allocataires d'un ensemble de communautés de communes. Les variables sont les suivantes :

**Ident** : Variable identificatrice des individus.

**ACT\_ALLOC, ACT\_CON** : Activité de l'allocataire (ACT\_ALLOC) et de son conjoint (ACT\_CON), les modalités sont les suivantes : "1-Actif", "2-Etudiant", "3-Chomeur", "4-Retraité", "5-Inactif", "6-Inconnu".

**age2006\_alloc, age2006\_conj** : Age, fin 2006, de l'allocataire et de son conjoint.

**sexe\_alloc** : Sexe de l'allocataire (intéressant uniquement pour les allocataires isolés).

**annnen1, annnen2,..., annnen7** : Année de naissance du i-ème enfant (pas de famille de plus de 7 enfants dans le fichier).

**RMI** : Identifie le versement du RMI (1 si versement, 0 sinon).

**dtdeMrmi** : Date de demande du RMI.

**C\_ss\_enf, C\_av\_enf, I\_ss\_enf, I\_av\_enf** : Variables identificatrices (valeurs 0 et 1) sur la situation familiale.

**Nbenf** : Nombre d'enfant(s) de l'allocataire.

#### Exercice 1 : Importation

Le fichier de données *caf2006.csv* se trouve sur le portail.

1. Importer les données en utilisant, dans une étape DATA, les instructions :  
informat dtdeMrmi ddmmyy10. ;  
format dtdeMrmi ddmmyy10. ;
2. Vérifier le résultat de l'importation par une PROC CONTENTS et une vue (PROC PRINT)
3. Que se passerait-il si l'on omet, dans l'étape DATA :
  - a. L'instruction "informat dtdeMrmi ddmmyy10.;" ? (Commenter le message d'erreur)
  - b. Uniquement l'instruction "format dtdeMrmi ddmmyy10.;" ?

#### Exercice 2 : Modification de la table

1. Créer les variables suivantes :
  - a. *Agenf1* : age du plus jeune enfant (en 2006).
  - b. *Sitfam* : situation familiale. Quatre modalités : "Couple sans enfant", "Couple avec enfant", "Isolé sans enfant", "Isolé avec enfant".
  - c. *Anc\_an* : ancienneté (en années) dans le RMI.

- d. *Anc\_mois* : ancienneté en mois dans le RMI (en décembre 2006).
2. Créer une table **rmi2006.sas7bdat** contenant uniquement les allocataires bénéficiaires du RMI.
3. Création de formats :
  - a. Créer un format "Homme/Femme" qui sera par la suite applicable à la variable sexe du jeu de données Caf2006.
  - b. Créer un format "Oui/Non" pour les variables identificatrices.
  - c. Créer, pour la variable *nbenf*, un format donnant les modalités : "Pas d'enfant", "1 enfant", "2 enfants au moins". Quel est l'intérêt d'utiliser ici un format plutôt qu'un changement de variable ?
  - d. Créer, pour la variable *sitfam*, un format donnant les modalités : "Couple", "Isolé".
4. Utilisation des formats :
  - a. Afficher la table **rmi2006.sas7bdat** en utilisant les formats précédents.
  - b. Affecter de façon permanente les formats précédents à la table **rmi2006.sas7bdat**.
  - c. Afficher à présent la table **rmi2006.sas7bdat** sans utiliser le format "Homme/Femme".

## Partie 2

### Exercice 1

On dispose d'un jeu de données concernant la production de pièces mécaniques. On dispose des dates de début de fabrication, ainsi que du nombre de jours passés par la pièce mécanique dans chacun des 4 ateliers de l'entreprise. On veut écrire un programme qui nous indique, pour chaque pièce, la date de sortie de l'atelier.

Saisissez les données suivantes dans une étape data

Produit	Début de fabrication	Atelier 1	Atelier 2	Atelier 3	Atelier 4
tal863510v	15/12/2011	5	8	11	4
tal863511v	17/01/2012	3	2	3	2
tal863512v	19/01/2012	8	12	8	12
tar248621d	22/01/2012	14	2	11	8

Ecrire le script permettant d'obtenir la date de sortie de l'atelier.

### Exercice 2 : L'instruction RETAIN (FACULTATIF)

- a. Créer une table SAS comportant une seule variable X et 7 individus prenant pour valeurs : 1, 5, 4, 2, 6, 9 et 8
- b. Définir une nouvelle variable Y1 contenant les sommes successives de X.
- c. Définir une nouvelle variable Y2 contenant les maximas successifs de X.
- d. Définir une nouvelle variable Y3 contenant les différences entre deux valeurs successives de X.

Note : On répondra aux questions b., c. et d. à l'aide de l'instruction RETAIN utilisée dans une étape DATA.