



Introduction à SAS

Atelier de formation 12, hiver 2016

Plateforme de recherche clinique et évaluative

Méthodologie et biostatistique

Anne-Sophie Julien

Anne-sophie.julien@crchudequebec.ulaval.ca

SAS version 9.3, SAS Institute, Cary, NC

Objectifs d'apprentissage

1. Connaître le fonctionnement de SAS
2. Apprendre les principes de base de programmation
3. Manipuler des fichiers de données
4. Créer des variables

Plan de la présentation

1. Interface
2. Principes de programmation
3. Importation de données
4. Formats
5. Étape DATA
6. Procédures de base
7. Énoncés globaux
8. Références
9. Exercices

Téléchargement de SAS

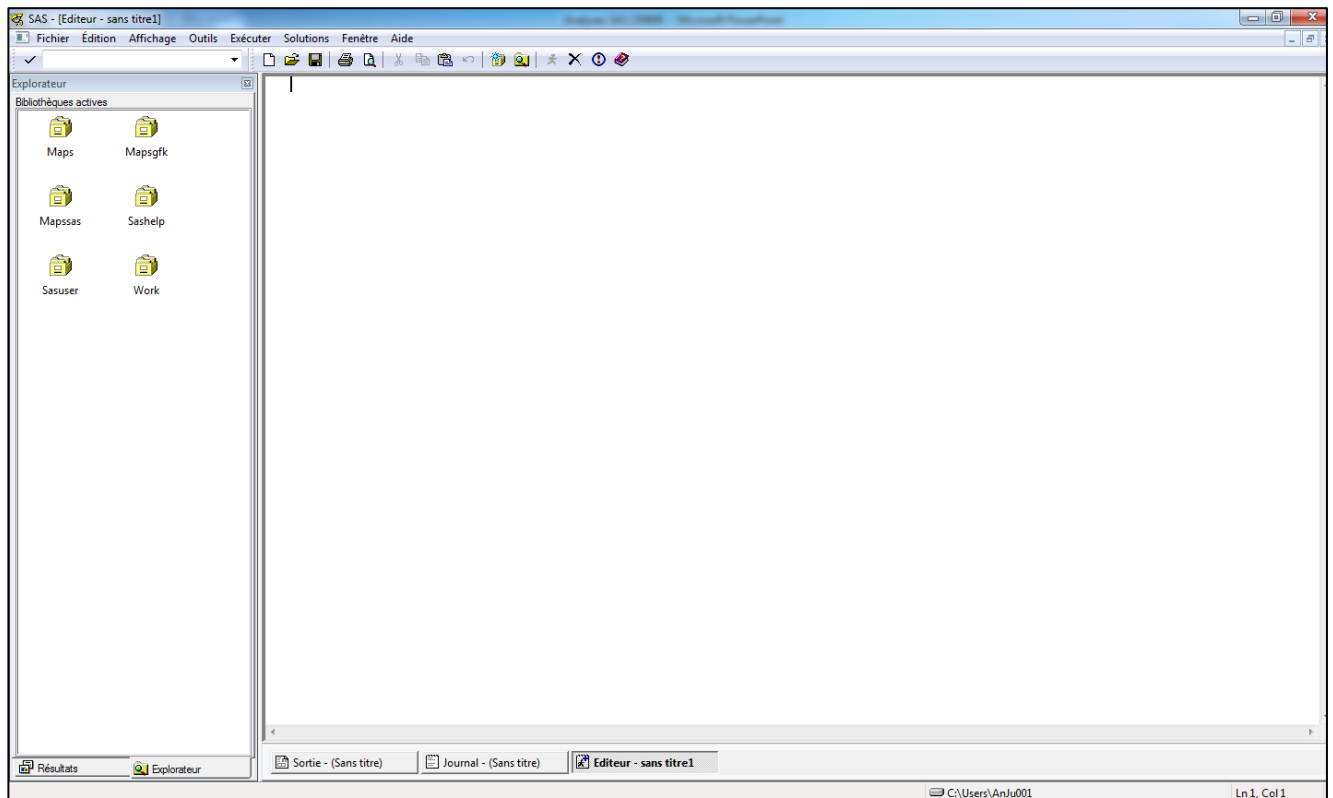
- Versions gratuites de SAS (limitées, avec interface différente de la version payante) pour étudiant et personnel universitaire, dans un but d'apprentissage :
 - SAS University Edition : http://www.sas.com/en_us/software/university-edition.html
 - Nécessite l'installation de VMWare Player et l'activation de la virtualisation dans votre BIOS si ce n'est pas déjà actif.
 - SAS On Demand : http://www.sas.com/en_us/industry/higher-education/on-demand-for-academics.html
 - Nécessite une connexion internet. Exécution et enregistrement des données sur les serveurs de SAS.
- Version payante de SAS pour tous, logiciel complet, version 9.4
 - License avec l'Université Laval (<https://www.dti.ulaval.ca/>)
 - 115\$, pour les étudiants, ou 230\$ pour les employés, jusqu'au 31 mai 2016
 - Nécessite Windows

Exemple de données

Contexte : Décès chez des patients atteints de cancer, suite à une chirurgie pour retirer la tumeur

Variable	Valeurs
Id	Numéro d'identification
Décès	0: Vivant, 1: Mort
Sexe	Femme, Homme
Fumeur	0: Non Fumeur, 1: Occasionnel, 2: Régulier
Diabète	Oui, Non
IMC	Nombre
Diamètre	Nombre (mm)
Date_nais	Date de naissance
Date_chir	Date de la chirurgie
Date_fin_suivi	Date de fin de suivi (décès ou dernière visite)
Hopital	Acronyme de l'hôpital et province

1. INTERFACE



- a. Éditeur (Editor): écriture de code par l'utilisateur
- b. Journal (Log): lorsque le code de l'utilisateur est exécuté, messages sur le bon ou mauvais fonctionnement des commandes soumises
- c. Sortie (Output): affichage des résultats
- d. Dossier en cours (Current Folder) : répertoire actif où les fichiers de résultats générés seront enregistrés, où des données seront importées
- e. Explorateur (Explorer):
 - i. Bibliothèque temporaire « Work ». Données effacées à la fermeture de SAS. Pour y faire référence : work.donnees (two-level name) ou donnees (one-level name)
 - ii. Bibliothèque permanente sur l'ordinateur. Données conservées sur l'ordinateur à la fermeture de SAS. Pour y faire référence : biblio.donnees
- f. Résultats (Results) : tables des matières des sorties / résultats.

g. Menus

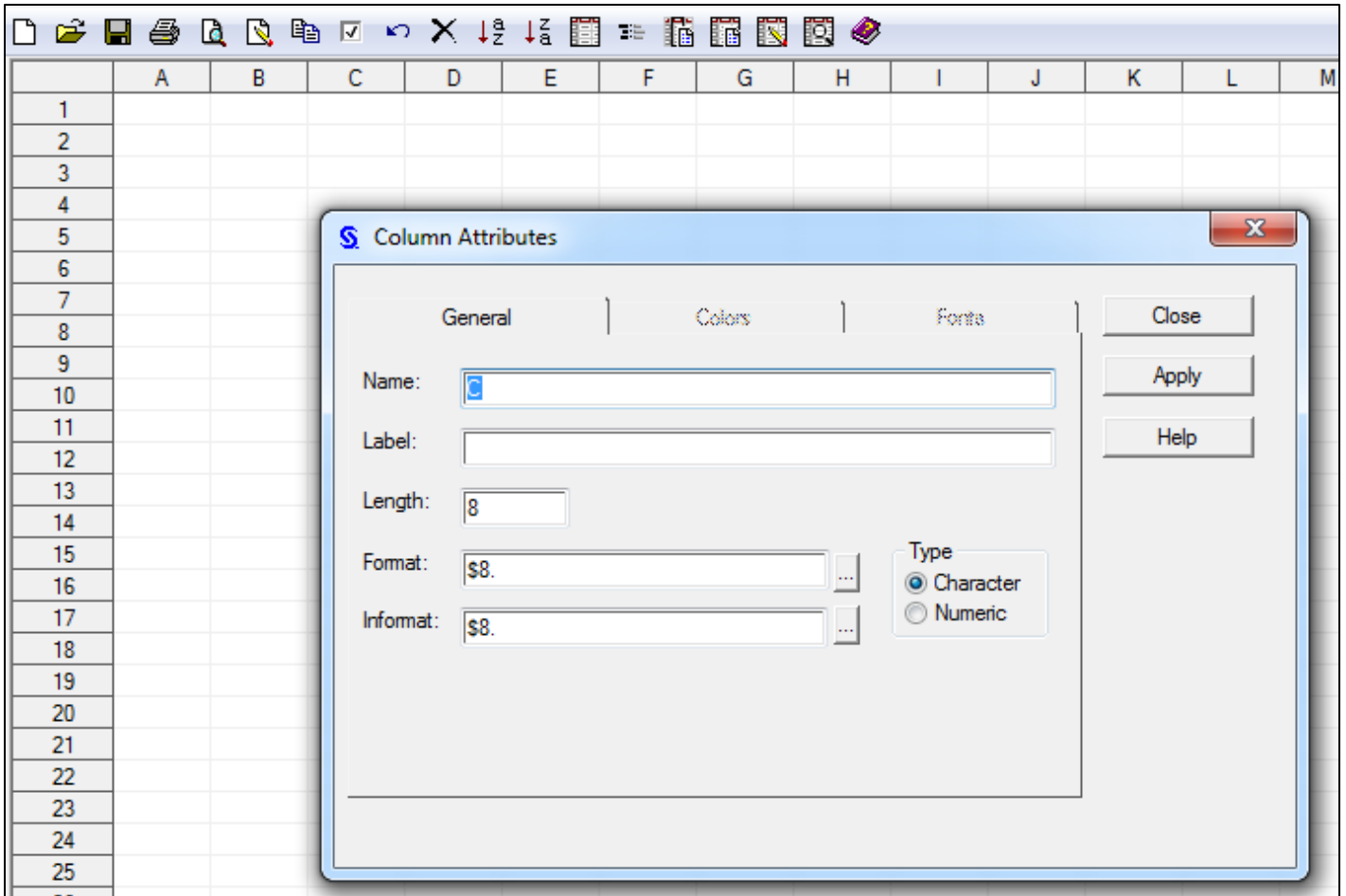
i. Barres d'outils de : Éditeur / Journal / Sortie



ii. Fichier / Édition /Affichage

iii. Outils (Tools)

1. Éditeurs de table SAS : saisie de données



2. Options -> Préférences

a. Résultats : Listing vs HTML

iv. Aide : intégrée vs en ligne. Syntaxe et exemples de toutes les fonctions, étapes et procédure. Recherche sur Google des messages d'erreurs, souvent des solutions proposées.

2. PRINCIPES DE PROGRAMMATION DANS L'ÉDITEUR

a. Couleurs

- i. Vert : commentaires de l'utilisateur
- ii. Bleu marin : code SAS; DATA et procédures
- iii. Noir : code de l'utilisateur
- iv. Bleu royal : code SAS; options et énoncés globaux
- v. Jaune : Données saisies directement dans le code
- vi. Mauve : entre guillemets; titres, notes de bas de page et étiquettes
- vii. Turquoise : formats et nombres
- viii. Rouge : code SAS non reconnu

b. Types d'énoncés

- i. Commentaires : texte de l'utilisateur entre les symboles /* et */ qui n'est pas lu par SAS
- ii. Étapes DATA : manipuler les données et variables, ne produisent pas de résultats en sortie
- iii. Procédures PROC : manipuler les fichiers de données et faire des analyses, générer des sorties
- iv. Énoncés globaux : options générales à l'extérieur d'un DATA ou PROC
- v. Fin d'un énoncé : point-virgule, RUN et QUIT. Procédure ouverte : "PROC / DATA step running"

c. Mise en page

- i. Indentation; indenter le cœur des procédures
- ii. Majuscules (code SAS), Minuscule (code de l'utilisateur)

```
DATA donnees_brutes;
    INPUT ancienne_variable;
DATALINES;
182
284
281
;
RUN;

/*Ceci est un commentaire*/

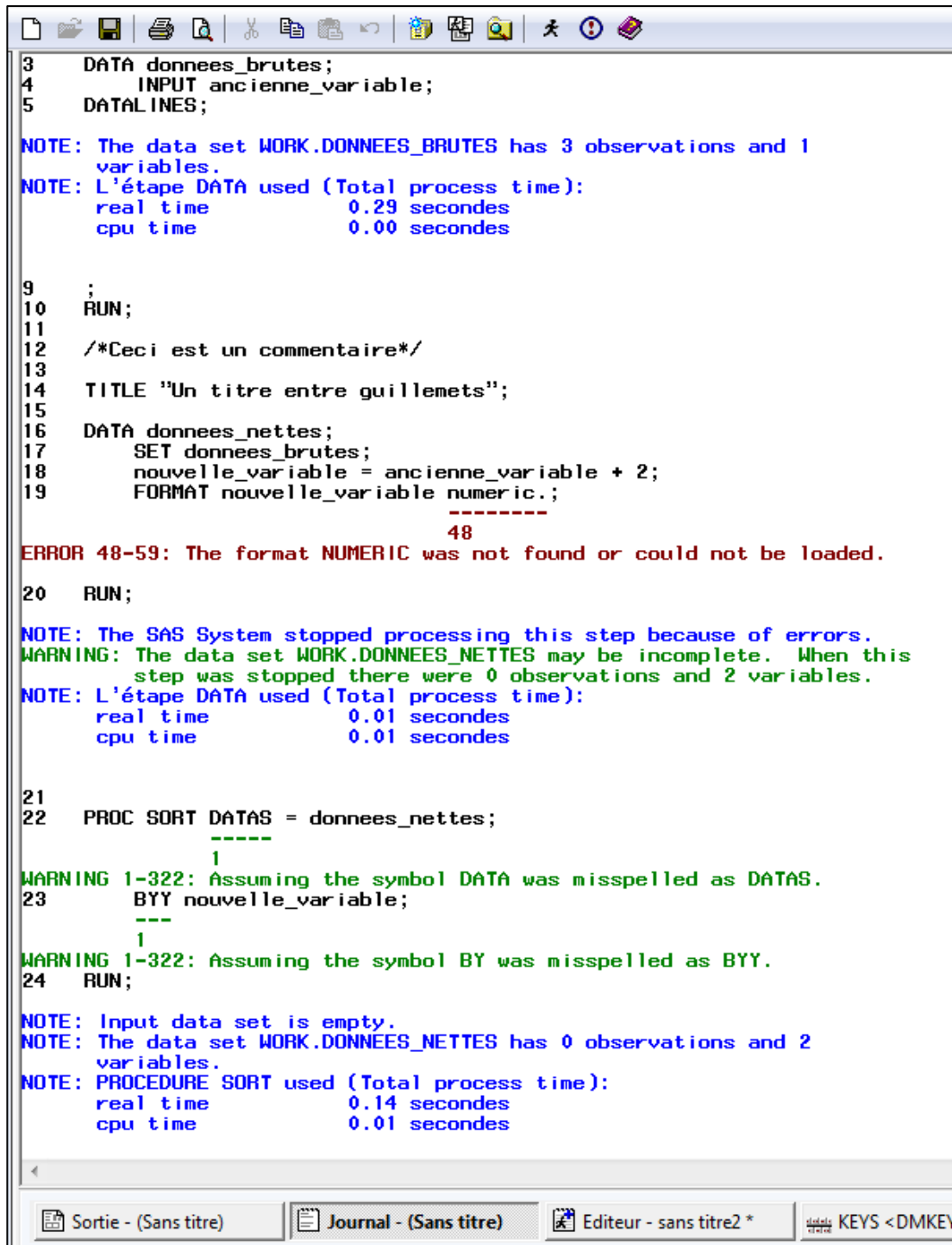
TITLE "Un titre entre guillemets";

DATA donnees_nettes;
    SET donnees_brutes;
    nouvelle_variable = ancienne_variable + 2;
    FORMAT nouvelle_variable numeric.;
RUN;

PROC SORT DATAS = donnees_nettes;
    BYY nouvelle_variable;
RUN;
```

d. Exécution

- i. Clic droit vs Menu
- ii. Vérification journal et résultat
 1. Noir : code soumis par l'utilisateur
 2. Bleu : message ok par SAS
 3. Vert : Warnings
 4. Rouge : Erreur dans le code
 - a. Syntaxe : options invalides, guillemets ouverts, point-virgule manquant, etc.
 - b. Logique : données ne se prêtent pas à ce qui est demandé



```
3  DATA donnees_brutes;
4  INPUT ancienne_variable;
5  DATALINES;

NOTE: The data set WORK.DONNEES_BRUTES has 3 observations and 1
variables.
NOTE: L'étape DATA used (Total process time):
      real time      0.29 secondes
      cpu time       0.00 secondes

9  ;
10 RUN;
11
12 /*Ceci est un commentaire*/
13
14 TITLE "Un titre entre guillemets";
15
16 DATA donnees_nettes;
17 SET donnees_brutes;
18 nouvelle_variable = ancienne_variable + 2;
19 FORMAT nouvelle_variable numeric.;
                                -----
                                48
ERROR 48-59: The format NUMERIC was not found or could not be loaded.

20 RUN;

NOTE: The SAS System stopped processing this step because of errors.
WARNING: The data set WORK.DONNEES_NETTES may be incomplete. When this
step was stopped there were 0 observations and 2 variables.
NOTE: L'étape DATA used (Total process time):
      real time      0.01 secondes
      cpu time       0.01 secondes

21
22 PROC SORT DATAS = donnees_nettes;
                                -----
                                1
WARNING 1-322: Assuming the symbol DATA was misspelled as DATAS.
23 BYY nouvelle_variable;
                                ---
                                1
WARNING 1-322: Assuming the symbol BY was misspelled as BYY.
24 RUN;

NOTE: Input data set is empty.
NOTE: The data set WORK.DONNEES_NETTES has 0 observations and 2
variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time      0.14 secondes
      cpu time       0.01 secondes
```

Sortie - (Sans titre) | Journal - (Sans titre) | Editeur - sans titre2 * | KEYS < DMKEY

3. IMPORTATION DE DONNÉES

- a. Types de fichier : Excel, Access, SPSS, Stata, CSV, txt...
- b. Problèmes courants
 - i. Le fichier importé ne doit pas avoir plus de 255 colonnes habituellement (dépend des versions).
 - ii. Noms des variables à renommer s'ils ne respectent pas les règles : maximum 32 caractères, doit commencer par une lettre ou « underscore », le reste peut aussi comprendre des chiffres.
 - iii. Variables numériques lues comme caractères si beaucoup de cellules vides au début ou si des caractères sont intégrés dans certaines cellules : les nettoyer au préalable.
 - iv. Attention aux points versus virgules utilisés comme des décimales; il faut utiliser le symbole par défaut du système. Sinon, les valeurs numériques seront lues en caractères.
 - v. Variables de dates lues comme caractères ou décalées de 60 ans (Excel compte à partir du 1^{er} janvier 1900, alors que SAS travaille à partir du 1^{er} janvier 1960) : uniformiser le format de date dans Excel.
 - vi. Valider l'importation en examinant le journal, le fichier créé dans l'explorateur, les « Column Attributes », ou en exécutant diverses procédures (par exemple : CONTENTS, FREQ, MEANS).
- c. Via Menu -> Fichier -> « Importer données »
 - i. Choix du type de fichier à importer
 - ii. Sélectionner le chemin du fichier à importer dans l'ordinateur
 - iii. Donner un nom au fichier à importer et sélectionner la bibliothèque dans laquelle l'enregistrer.

d. Avec la procédure IMPORT

```

/*!!! Attention !!!
   Le DBMS à saisir change souvent selon la version de SAS, Windows et Excel
   utilisée
   SHEET peut être remplacé par RANGE selon les versions*/
PROC IMPORT DATAFILE = ".\Atelier SAS Données.xlsx"
  OUT = donneesB
  DBMS = excel
  REPLACE;
  SHEET = "Données$B2:L102";
RUN;

/*Voir les caractéristiques des variables et le nombre d'observations des données*/
PROC CONTENTS DATA = donneesB;
RUN;

```

```

141 PROC IMPORT DATAFILE = ".\Atelier SAS Données.xlsx"
142     OUT = donneesB
143     DBMS = excel
144     REPLACE;
145     SHEET = "Données$B2:L102";
146 RUN;

NOTE: WORK.DONNEESB data set was successfully created.
NOTE: The data set WORK.DONNEESB has 100 observations and 11 variables.
NOTE: PROCEDURE IMPORT used (Total process time):
      real time          0.64 secondes
      cpu time           0.21 secondes

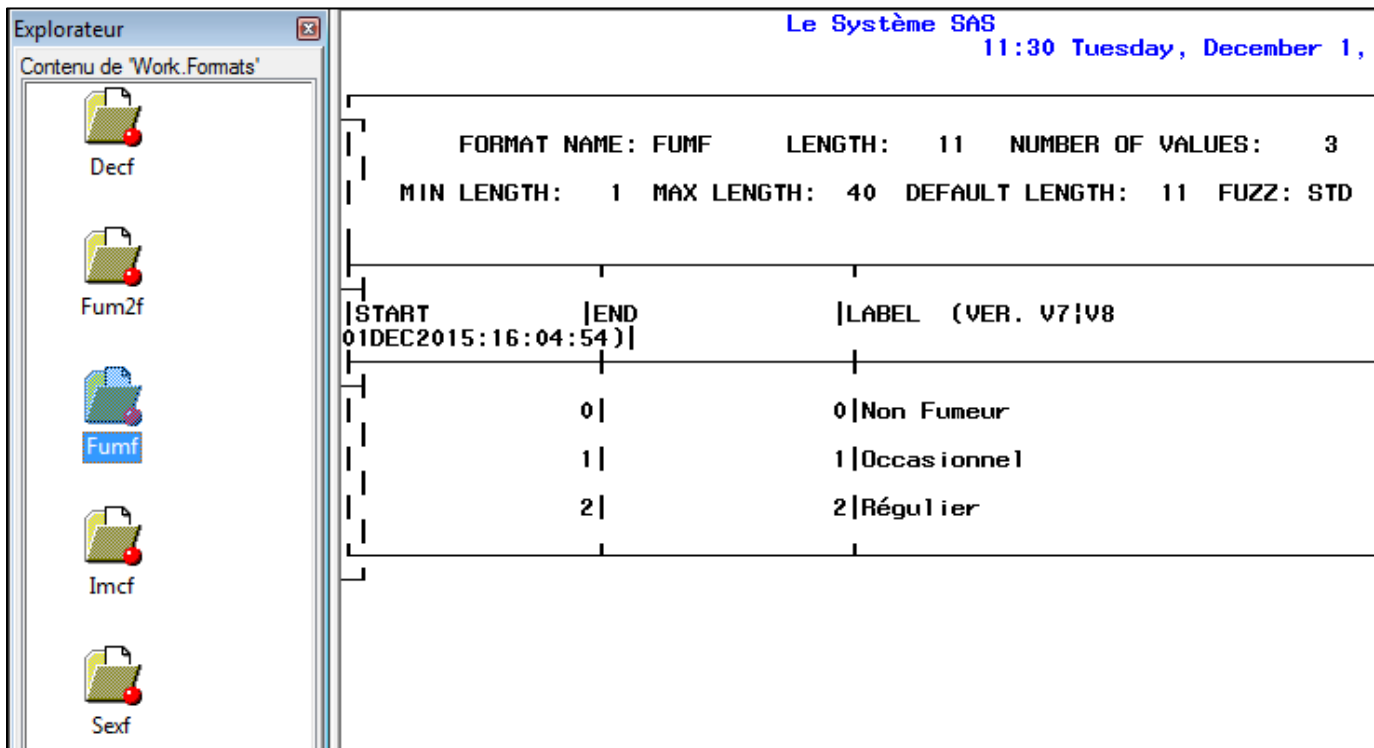
```

Liste alphabétique des variables et des attributs						
N	Variable	Type	Long.	Format	Informat	Libellé
2	Deces	Num.	8			Deces
5	Diabete	Texte	3	\$3.	\$3.	Diabete
7	Diametre	Num.	8			Diametre
4	Fumeur	Num.	8			Fumeur
11	Hopital	Texte	14	\$14.	\$14.	Hopital
6	IMC	Num.	8			IMC
1	Id	Num.	8			Id
3	Sexe	Texte	5	\$5.	\$5.	Sexe
9	date_chir	Num.	8	DATE9.	DATE9.	date_chir
10	date_fin_suivi	Num.	8	DATE9.	DATE9.	date_fin_suivi
8	date_nais	Num.	8	DATE9.	DATE9.	date_nais

4. FORMATS

- Énoncé VALUE pour spécifier les étiquettes de chaque valeur
- Nom du format ne doit pas finir par un chiffre. Commencer par un \$ pour les variables caractères.
- Options pour certaines valeurs : HIGH, LOW, OTHER. Utiliser < avec le tiret pour exclure une valeur d'une étendue.

```
PROC FORMAT LIBRARY = WORK;  
  VALUE decf 1 = "Mort" 0 = "Vivant" OTHER = "!ERREUR!";  
  VALUE $sexf "Femme" = "Féminin" "Homme" = "Masculin";  
  VALUE fumf 0 = "Non Fumeur" 1 = "Occasionnel" 2 = "Régulier";  
  VALUE fum2f 0 = "Non Fumeur" 1 = "Fumeur";  
  VALUE imcf low-<18 = "Insuffisance" 18-25 = "Normal" 25<-high = "Surplus";  
RUN;
```



The screenshot shows the SAS Explorer window on the left, displaying the contents of the 'Work.Formats' library. The files listed are Decf, Fum2f, Fumf (highlighted), Imcf, and Sexf. The main window shows the SAS output for the PROC FORMAT statement, including the format name, length, number of values, and a table of values and labels.

Le Système SAS
11:30 Tuesday, December 1,

FORMAT NAME: FUMF LENGTH: 11 NUMBER OF VALUES: 3
MIN LENGTH: 1 MAX LENGTH: 40 DEFAULT LENGTH: 11 FUZZ: STD

START	END	LABEL (VER. V7 V8)
0	0	Non Fumeur
1	1	Occas ionnel
2	2	Régul ier

5. ÉTAPE DATA

- a. Création d'un nouveau jeu de données à partir de données existantes

```
DATA donnees99;  
    SET donneesB;
```

- b. Conservation d'un sous-échantillon

```
/*Conserver seulement un sous échantillon  
. = donnée manquante pour les variables numériques  
Autres syntaxes équivalentes:  
IF id ne 100;  
IF id not in (.,100) ;  
IF id > 99 THEN DELETE;*/  
WHERE . < id <= 99;
```

```
NOTE: There were 99 observations read from the data set WORK.DONNEESB.  
      WHERE (id). and id<=99);  
NOTE: The data set WORK.DONNEES99 has 99 observations and 33 variables.  
NOTE: L'étape DATA used (Total process time):  
      real time          0.07 secondes  
      cpu time           0.04 secondes
```

- c. Correction conditionnelle de valeurs numériques et alphanumériques

```
/*Corrections de valeurs pour un individu  
Vérifier une condition, puis effectuer un bloc d'opérations ou une seule  
opération  
Vide ou "" = donnée manquante pour les variables caractères*/  
IF id = 56 THEN diabete = "Oui";  
IF id = 55 THEN DO;  
    imc = 23.6;  
    sexe = "";  
END;
```

donneesB				
Obs.	Id	Diabete	IMC	Sexe
55	55	Non	26.3	Homme
56	56	Non	23.3	Homme

donnees99				
Obs.	no	Diabete	IMC	Sexe
55	55	Non	23.6	
56	56	Oui	23.3	Homme

d. Recodification de variables nominales

```

/*Recoder des variables*/
IF fumeur = 0 THEN fumeur2 = 0;
ELSE IF fumeur = 1 THEN fumeur2 = 1;
ELSE IF fumeur = 2 THEN fumeur2 = 1;
/*Syntaxes équivalentes pour les 2 ELSE ci-dessus:
  -ELSE IF 1 <= fumeur <= 2 THEN fumeur2 = 1;
  -ELSE IF fumeur in (1,2) THEN fumeur2 = 1;
  -ELSE IF fumeur > 0 THEN fumeur2 = 1;*/
ELSE fumeur2 = .; /*facultatif ici: variables non assignées sont mises à
                    manquantes*/

```

Obs.	Fumeur	fumeur2
93	Non Fumeur	Non Fumeur
94	Non Fumeur	Non Fumeur
95	Occasionnel	Fumeur
96	Occasionnel	Fumeur
97	Régulier	Fumeur

e. Opérations sur des variables numériques

```

/*Opérations sur variables numériques*/
diametreR = ROUND(diametre,1); /*Arrondir à l'unité*/
diametreI = INT(diametre); /*Garder la partie entière, Integer*/
diam_cm = diametre /10; /*Opérateurs mathématiques : * / - + */
volume = CONSTANT('pi')*(diametre**3)/6; /*volume d'une sphère (pi*diam^3/6).
                                           **3 = exposant 3*/
sqrt_volume = SQRT(volume); /*racine carrée*/
ln_volume = LOG(volume); /*logarithme népérien*/
nombre_bidon = MEAN(OF volume,imc); /*moyenne. Aussi disponible: SUM*/

```

Obs.	Diametre	diametreR	diametreI	diam_cm	volume	sqrt_volume	ln_volume	IMC	nombre_bidon
1	1.6	2	1	0.16	2.1447	1.46447	0.76298	27.5	14.822330292
2	3.2	3	3	0.32	17.1573	4.14214	2.84242	19.2	18.178642339
3	3.5	4	3	0.35	22.4493	4.73807	3.11126	21.6	22.024648752
4	4.4	4	4	0.44	44.6022	6.67849	3.79778	22.1	33.35111905
5	4.5	5	4	0.45	47.7129	6.90746	3.86520	24.4	36.056469213

f. Fonctions sur des dates

```

/*Variables de dates (variables numériques)*/
IF date_chir >= '01JAN2000'd THEN methode_chir = "Nouvelle";
ELSE methode_chir = "Ancienne"; /*identifier des valeurs de dates: "Date Constant"*/
Year_chir = YEAR(date_chir); /*extraire l'année d'une date.
                                Aussi disponible: MONTH, WEEK, WEEKDAY, QTR, DAY*/
j = 23;
m = 11;
a = 2014;
new_date = MDY(m,j,a); /*créer 1 variable de date unique à partir de 3 variables*/

age_chir = YRDIF(date_nais,date_chir,'AGE'); /*calculer la différence entre 2 dates
                                             en années*/
suivi_ans = YRDIF(date_chir,date_fin_suivi,'ACT/ACT'); /*calculer la différence
                                                       entre 2 dates en années, en divisant le nombre actuel de jours de
                                                       la période par le nombre actuel de jours de chaque année*/
suivi_weeks = (date_fin_suivi - date_chir)/7; /*différence de deux dates = résultat
                                             en jours*/

```

donneesB

donnees99

Obs.	date_chir	date_nais	date_fin_suivi
21	18OCT2001	07JAN1983	05FEB2004
22	21JUL2000	14OCT1981	27DEC2004
23	24APR1999	21JUL1980	27FEB2004
24	25JAN1998	28APR1979	15MAR2000
25	28OCT1996	02FEB1978	12DEC1999

Obs.	methode_chir	Year_chir	new_date	age_chir	suivi_ans	suivi_weeks
21	Nouvelle	2001	23NOV2014	18.7781	2.30111	120.000
22	Nouvelle	2000	23NOV2014	18.7671	4.43443	231.429
23	Ancienne	1999	23NOV2014	18.7589	4.84615	252.857
24	Ancienne	1998	23NOV2014	18.7452	2.13643	111.429
25	Ancienne	1996	23NOV2014	18.7342	3.12280	162.857

g. Manipulation de variables alphanumériques

```

/*Manipulations de caractères*/
LENGTH province $10; /*Limiter la longueur de la nouvelle variable province à 10
                    caractères. (Longueur de SCAN par défaut = 200)*/
hopitalCorr = TRANWRD(hopital,"Prairies","Alberta"); /*Remplacer Prairies par
                                                       Alberta dans hôpital*/
province = SCAN(hopitalCorr,2,','); /*Extraire la 2e partie de hopital, après la
                                   virgule. Plusieurs délimiteurs possibles.
                                   Position négative pour partir de la fin*/
centre = SUBSTR(hopital,1,4); /*Extraire les 4 premiers caractères de hopital*/
hopitalLong = "Hôpital " || centre || "de la province de " || province;
/*Concaténer du texte et des variables*/
centre_minus = LOWCASE(centre); /*Aussi disponibles: UPCASE et PROPCASE*/
IF FIND(hopital,"CHU") > 0 THEN centre_hosp_univ = 1;
ELSE centre_hosp_univ = 0; /*Identifier les patients avec CHU dans le nom de
                             l'hôpital*/

```

Obs.	Hopital	hopitalCorr	province	centre	hopitalLong	centre_minus	centre_hosp_univ
1	ECHC, Prairies	ECHC, Alberta	Alberta	ECHC	Hôpital ECHC de la province de Alberta	echc	0
2	CHUS, Québec	CHUS, Québec	Québec	CHUS	Hôpital CHUS de la province de Québec	chus	1
3	ECHC, Prairies	ECHC, Alberta	Alberta	ECHC	Hôpital ECHC de la province de Alberta	echc	0
4	ECHC, Prairies	ECHC, Alberta	Alberta	ECHC	Hôpital ECHC de la province de Alberta	echc	0
5	CHUL, Québec	CHUL, Québec	Québec	CHUL	Hôpital CHUL de la province de Québec	chul	1

h. Opérations sur plusieurs variables simultanément

```
/*Effectuer une même opération sur plusieurs variables simultanément
   Toutes les valeurs 999 des variables mentionnées dans l'ARRAY toto seront
   recodées à des valeurs manquantes automatiquement*/
ARRAY toto imc diametre;
DO OVER toto;
   IF toto = 999 THEN toto = .;
END;
```

i. Assignation de formats permanents prédéfinis et définis par l'utilisateur

```
FORMAT /*Définis par l'utilisateur*/
   deces decf. sexe sexf. fumeur fumf. fumeur2 fum2f.
   /*Définis par SAS*/
   date_nais DDMMYY10. /*ex.: 01/01/2016*/ date_chir YEAR4. /*ex.: 2016*/
   date_fin_suivi MMYYS5. /*ex.: 01/16*/ new_date DATE9. /*ex.: 01JAN2016*/
   nombre_bidon 6.0 diabete $1.;
```

j. Conserver un sous-ensemble de variables

```
/*Pour garder seulement un sous-ensemble de variables
   Deux commandes équivalentes ici
   Équivalent à "DATA donnees99 (DROP = j m a);"
   ou "KEEP la liste de toutes les variables conservées"; */
DROP j m a ;
```

k. Renommer des variables avec RENAME

```
/*Renommer des variables*/
RENAME id = no;
```

l. Créer des étiquettes pour le nom des variables avec LABEL

```
/*Ajouter des étiquettes pour remplacer les noms des variables lors de l'affichage
   des résultats*/
LABEL diametre = "Diamètre (mm)"
   diam_cm = "Diamètre (cm)"
   deces = "Décès";
```

```
RUN;
```

m. Concaténation de fichiers avec SET

```

/*Importer des données supplémentaires à ajouter aux données initiales*/
PROC IMPORT DATAFILE = ".\Atelier SAS Données.xlsx"
  OUT = new_obs
  DBMS = excel
  REPLACE;
  SHEET = "new_obs";
RUN;
PROC IMPORT DATAFILE = ".\ Atelier SAS Données.xlsx"
  OUT = new_var
  DBMS = excel
  REPLACE;
  SHEET = "new_var";
RUN;

/*Concaténer deux fichiers de données*/
DATA donneesB_new_obs;
  SET donneesB new_obs;
RUN;

```

	Id	Deces	Sexe	Fumeur	Diabete	IMC	Diametre	date_nais
99	99	1	Homme	2	Non	24.7	19.4	28AUG1960
100	100	1	Homme	2	Non	19.9	27.8	05JUN1959
101	101	0		.	Non	.	.	.
102	102	0		.	Oui	.	.	.
103	103	0		.	Non	.	.	.
104	104	1		.	Oui	.	.	.
105	105	1		.	Non	.	.	.

n. Juxtaposition de fichiers de données avec MERGE selon une variable clé

```

/*Juxtaper deux fichiers de données
  Trier les observations au préalable selon la clé de tri avec SORT*/

PROC SORT DATA = donneesB;
  BY id;
RUN;

PROC SORT DATA = new_var;
  BY id;
RUN;

DATA donnees_new_var;
  MERGE donneesB new_var;
  BY id;
RUN;

```

Id	Décès	Rythme	gsanguin
1	Vivant	.	.
2	Vivant	.	.
3	Vivant	.	.
4	Vivant	.	.
5	Vivant	.	.
6	Vivant	.	.
7	Vivant	103	O-

6. PROCÉDURES (PROC) DE BASE

- Importer les données d'un fichier externe (voir PROC IMPORT dans la section 3)
- Connaitre le contenu d'un jeu de données (voir exemple de résultat à la section 3)

```
/*Connaitre le contenu d'un jeu de données*/
PROC CONTENTS DATA = donnees99 VARNUM;
RUN;
```

- Afficher le contenu complet ou partiel d'une banque de données (voir résultats dans la section 5)

```
PROC PRINT DATA = donneesb (FIRSTOBS = 55 OBS = 56);
    VAR id diabete imc sexe;
RUN;
PROC PRINT DATA = donnees99 (FIRSTOBS = 55 OBS = 56);
    VAR no diabete imc sexe;
    FORMAT Diabete $3. Sexe $5. ;
RUN;
PROC PRINT DATA = donnees99 (FIRSTOBS = 93 OBS = 97);
    VAR fumeur fumeur2;
RUN;
PROC PRINT DATA = donnees99 (OBS = 5);
    VAR diametre diametreR diametreI diam_cm volume sqrt_volume ln_volume imc
        nombre_bidon;
    FORMAT nombre_bidon best12.;
RUN;
PROC PRINT DATA = donneesB (FIRSTOBS = 21 OBS = 25);
    VAR date_chir date_nais date_fin_suivi;
RUN;
PROC PRINT DATA = donnees99 (FIRSTOBS = 21 OBS = 25);
    VAR methode_chir year_chir /*j m a*/ new_date age_chir suivi_ans
        suivi_weeks;
RUN;
PROC PRINT DATA = donnees99 (OBS = 5);
    VAR hopital hopitalCorr province centre hopitalLong centre_minus
        centre_hosp_univ;
RUN;
```

- Trier les observations selon des variables

```
/*Trier les observations d'une banque de données*/
PROC SORT DATA = new_var OUT = new_var_tri;
    BY DESCENDING rythme;
RUN;
```

new_var

	id	Rythme	gsanguin
1	7	103	O-
2	16	89	AB+
3	34	115	B-
4	56	129	A+
5	75	97	O+

new_var_tri

	id	Rythme	gsanguin
1	56	129	A+
2	34	115	B-
3	7	103	O-
4	75	97	O+
5	16	89	AB+

e. Obtenir des statistiques descriptives pour valider les valeurs de variables quantitatives

```
/*Obtenir des statistiques descriptives pour des variables quantitatives*/
PROC MEANS DATA = donnees99;
    VAR diametre imc age_chir;
    CLASS deces;
RUN;
```

Décès	N Obs	Variable	Libellé	N	Moyenne	Ecart-type	Minimum	Maximum
Vivant	79	Diametre	Diamètre (mm)	79	8.91	5.70	1.20	26.00
		IMC	IMC	79	27.03	6.14	17.30	48.20
		age_chir		79	33.44	10.73	18.62	47.04
Mort	20	Diametre	Diamètre (mm)	20	13.23	6.62	2.70	26.30
		IMC	IMC	20	27.66	5.13	19.90	40.60
		age_chir		20	32.02	2.90	27.55	37.27

f. Générer des tableaux de fréquences simples ou croisés pour valider des variables qualitatives

```
/*Générer des tableaux de fréquences simples ou croisés*/
PROC FREQ DATA = donnees99;
    TABLES hopital fumeur*fumeur2;
    TABLES hopital*province /LIST MISSING;
RUN;
```

Hopital	Fréquence	Pourcentage	Fréquence cumulée	Pctage. cumulé
CHEO, Ontario	6	6.06	6	6.06
CHUL, Québec	6	6.06	12	12.12
CHUM, Québec	14	14.14	26	26.26
CHUS, Québec	9	9.09	35	35.35
CMCH, Ontario	13	13.13	48	48.48
ECHC, Prairies	18	18.18	66	66.67
HSFA, Québec	23	23.23	89	89.90
TBCC, Prairies	10	10.10	99	100.00

Fréquence Pourcentage Pctage en ligne Pctage en col.	Table de Fumeur par fumeur2			
	Fumeur(Fumeur)	fumeur2		
		Non Fumeur	Fumeur	Total
	Non Fumeur	80	0	80
80.81		0.00	80.81	
100.00		0.00		
100.00		0.00		
Occasionnel	0	11	11	
	0.00	11.11	11.11	
	0.00	100.00		
	0.00	57.89		
Régulier	0	8	8	
	0.00	8.08	8.08	
	0.00	100.00		
	0.00	42.11		
Total	80	19	99	
	80.81	19.19	100.00	

Hopital	province	Fréquence	Pourcentage	Fréquence cumulée	Pctage. cumulé
CHEO, Ontario	Ontario	6	6.06	6	6.06
CHUL, Québec	Québec	6	6.06	12	12.12
CHUM, Québec	Québec	14	14.14	26	26.26
CHUS, Québec	Québec	9	9.09	35	35.35
CMCH, Ontario	Ontario	13	13.13	48	48.48
ECHC, Prairies	Alberta	18	18.18	66	66.67
HSFA, Québec	Québec	23	23.23	89	89.90
TBCC, Prairies	Alberta	10	10.10	99	100.00

7. ÉNONCÉS GLOBAUX

Énoncés en dehors d'un DATA ou PROC, pour contrôler des aspects plus généraux.

a. Pour définir une librairie pour enregistrer ou lire des fichiers : LIBNAME

```
/*Redéfinir la librairie sur l'ordinateur*/
LIBNAME myfiles "P:\Plateformes\PRC\Consult_ASJulien\Atelier 12 Intro SAS\";

DATA myfiles.donnees99;
    SET donnees99; /*équivalent à work.donnees99*/
RUN;
```

```
233 /*Redéfinir la librairie sur l'ordinateur*/
234 LIBNAME myfiles "P:\Plateformes\PRC\Consult_ASJulien\Atelier 12
234! Intro SAS\";
NOTE: Libref MYFILES was successfully assigned as follows:
      Engine:          V9
      Physical Name: P:\Plateformes\PRC\Consult_ASJulien\Atelier 12
      Intro SAS
235
236 DATA myfiles.donnees99;
237     SET donnees99; /*équivalent à work.donnees99*/
238 RUN;

NOTE: There were 99 observations read from the data set WORK.DONNEES99.
NOTE: The data set MYFILES.DONNEES99 has 99 observations and 33
      variables.
NOTE: L'étape DATA used (Total process time):
      real time          0.04 secondes
      cpu time           0.01 secondes
```

b. Pour ajouter un titre ou note de bas de pages : TITLE, FOOTNOTES

```
/*Mettre des titres et notes de bas de page*/
TITLE "Statistiques descriptives";
TITLE2 "Variables quantitatives";
FOOTNOTE "Tableau 1";
PROC MEANS DATA = donnees99 MAXDEC = 2;
    VAR diametre imc age_chir;
    CLASS deces;

RUN;

TITLE;
FOOTNOTE;
```

Statistiques descriptives Variables quantitatives								
Procédure MEANS								
Décès	N Obs	Variable	Libellé	N	Moyenne	Ecart-type	Minimum	Maximum
Vivant	79	Diametre	Diamètre (mm)	79	8.91	5.70	1.20	26.00
		IMC	IMC	79	27.03	6.14	17.30	48.20
		age_chir		79	33.44	10.73	18.62	47.04
Mort	20	Diametre	Diamètre (mm)	20	13.23	6.62	2.70	26.30
		IMC	IMC	20	27.66	5.13	19.90	40.60
		age_chir		20	32.02	2.90	27.55	37.27

Tableau 1

c. Pour contrôler les options générales de mise en page : OPTIONS

```
/*Contrôler les options générales de mise en page*/  
OPTIONS NODATE NUMBER ORIENTATION = PORTRAIT;
```

d. Pour enregistrer les résultats dans un fichier externe (RTF, HTML, PDF, DOC...);

```
/*Enregistrer les résultats dans un fichier RTF  
Aussi disponibles: PDF, Excel, DOC, HTML...*/
```

```
ODS RTF FILE = "./Résultats.rtf";
```

```
PROC MEANS DATA = donnees99 MAXDEC = 2;  
VAR diametre imc age_chir;  
CLASS deces;
```

```
RUN;
```

```
ODS RTF CLOSE;
```

Décès	Obs	Variable	Libellé	N	Moyenne	Ecart-type	Minimum	Maximum
Vivant	79	Diametre	Diamètre (mm)	79	8.91	5.70	1.20	26.00
		IMC	IMC	79	27.03	6.14	17.30	48.20
		age_chir		79	33.44	10.73	18.62	47.04
Mort	20	Diametre	Diamètre (mm)	20	13.23	6.62	2.70	26.30
		IMC	IMC	20	27.66	5.13	19.90	40.60
		age_chir		20	32.02	2.90	27.55	37.27

8. RÉFÉRENCES

- Documentation en ligne sur chacune des procédures, fonctions et création de DATA (syntaxe, exemples complets) : <http://support.sas.com/documentation/>
- Livre sur la programmation de base: SAS Certification Prep Guide : Base Programming for SAS 9, 3rd edition. SAS Institute Inc. : Cary, NC. 802 pages.
- Google : Très pratique pour rechercher des options ou messages d'erreur particuliers.
- Delwiche, L.D. et S.J. Slaughter (2012). The Little SAS Book : A Primer, 5th edition. SAS Institute Inc. : Cary, NC. 376 pages.
- Cody, R. (2015). An Introduction to SAS University Edition. SAS Institute Inc. : Cary, NC. 366 pages.