



Introduction à SAS - Atelier de formation 12, hiver 2016
Plateforme de recherche clinique et évaluative - Méthodologie et biostatistique
Anne-Sophie Julien - SAS version 9.3, SAS Institute, Cary, NC

Exercice

1. Compléter le code nécessaire pour générer le fichier de données « Nouvelle » à partir de l'onglet « Exercice » du fichier de données Excel « Atelier SAS Données.xlsx », ainsi que certains résultats. Le fichier créé devra être disponible sur l'ordinateur après la fermeture de SAS

Vue de « Exercice »

	Bebe	NAM	Poids	Grandeur	age_gest_jrs	Coord_mere	Hopital	An_nais	Mois_nais	Jour_nais
1	1	ABOF130608F	8.6	51	281	354, avenue du Tilleul, St-ClinClin	CHUM	13	6	8
2	2	SIMD140531G	7.4	48	264	23, rue du Bouleau, St-Henri	HDL	14	5	31
3	3	TURS131013G	6.5	49	254	934A, boulevard du Tremble, St-Sévérin	CHAUR	13	10	13
4	4	LAUJ110325G	9.7	56	273	2800, rang du Sapin, St-Augustin	CHUL	11	3	25
5	5	BARR090425F	4.1	31	196	92, ruelle du Merisier, St-Émile	CHUL	9	4	25
6	6	BOUM080705F	8.1	69	267	843, rue du Cèdre, Ste-Catherine	CHUL	8	7	5
7	7	MOIJ050825F	10.6	75	259	382, rang de l'Olivier, St-Raphaël	HDL	5	8	25
8	8	JULM010525F	11.4	64	269	29146, avenue du Palmier, Ste-Julie	HSCM	1	5	25
9	9	POUI141103H	8.5	58	251	2842, boulevard du Cocotier, St-Gabriel	HSFA	14	11	3
10	10	JULA130415F	7.9	42	270	1, rue de l'Épinette, St-Zachary	HSCM	13	4	15

Vue de « Nouvelle »

Bebe	Poids	Grandeur	Âge gestationnel (jrs)	Hopital	Mois_nais	region	imc	date_naissance	date_conception	ville	sexe
1	8.6	51	281	CHUM	Jan à Juin	Ailleurs	15.029	06/08/13	08/31/12	St-ClinClin	Fille
2	7.4	48	264	HDL	Jan à Juin	Capitale nationale	14.599	05/31/14	09/09/13	St-Henri	Garçon
3	6.5	49	254	CHAUR	Juil à Déc	Ailleurs	12.305	10/13/13	02/01/13	St-Sévérin	Garçon
4	9.7	56	273	CHUL	Jan à Juin	Capitale nationale	14.06	03/25/11	06/25/10	St-Augustin	Garçon
5	4.1	31	196	CHUL	Jan à Juin	Capitale nationale	19.393	04/25/09	10/11/08	St-Émile	Fille
6	8.1	69	267	CHUL	Juil à Déc	Capitale nationale	7.7333	07/05/08	10/12/07	Ste-Catherine	Fille
7	10.6	75	259	HDL	Juil à Déc	Capitale nationale	8.5657	08/25/05	12/09/04	St-Raphaël	Fille
8	11.4	64	269	HSCM	Jan à Juin	Ailleurs	12.651	05/25/01	08/29/00	Ste-Julie	Fille
9	8.5	58	251	HSFA	Juil à Déc	Capitale nationale	11.485	11/03/14	02/25/14	St-Gabriel	Garçon
10	7.9	42	270	HSCM	Jan à Juin	Ailleurs	20.357	04/15/13	07/19/12	St-Zachary	Fille

Sorties du Result Viewer

Variables par ordre de création						
N	Variable	Type	Long.	Format	Informat	Libellé
1	Bebe	Num.	8			Bebe
2	NAM	Texte	11	\$11.	\$11.	NAM
3	Poids	Num.	8			Poids
4	Grandeur	Num.	8			Grandeur
5	age_gest_jrs	Num.	8			age_gest_jrs
6	Coord_mere	Texte	39	\$39.	\$39.	Coord_mere
7	Hopital	Texte	5	\$5.	\$5.	Hopital
8	An_nais	Num.	8			An_nais
9	Mois_nais	Num.	8			Mois_nais
10	Jour_nais	Num.	8			Jour_nais

Moyenne et compagnie

Procédure MEANS

Variable	Libellé	N	Moyenne	Ecart-type	Minimum	Maximum
imc		10	13.618	4.073	7.733	20.357
Poids	Poids	10	8.280	2.078	4.100	11.400
taille	Grandeur	10	54.300	13.030	31.000	75.000
age_gest_jrs	Âge gestationnel (jrs)	10	258.400	23.675	196.000	281.000

Stats descriptives

Fréquences

Procédure FREQ

ville	Fréquence	Pourcentage	Fréquence cumulée	Pctage. cumulé
St-Augustin	1	10.00	1	10.00
St-ClinClin	1	10.00	2	20.00
St-Gabriel	1	10.00	3	30.00
St-Henri	1	10.00	4	40.00
St-Raphaël	1	10.00	5	50.00
St-Sévérin	1	10.00	6	60.00
St-Zachary	1	10.00	7	70.00
St-Émile	1	10.00	8	80.00
Ste-Catherine	1	10.00	9	90.00
Ste-Julie	1	10.00	10	100.00

Fréquence

Pourcentage

Pctage en ligne

Pctage en col.

Table de sexe par Mois_nais

sexe	Mois_nais(Mois_nais)		
	Jan à Juin	Juil à Déc	Total
Fille	4	2	6
	40.00	20.00	60.00
	66.67	33.33	
	66.67	50.00	
Garçon	2	2	4
	20.00	20.00	40.00
	50.00	50.00	
	33.33	50.00	
Total	6	4	10
	60.00	40.00	100.00

Code à compléter à partir des informations des deux pages précédentes :

```
PROC IMPORT DATAFILE = " _____ "
  OUT = _____
  DBMS = _____
  REPLACE;
  SHEET = _____
RUN;
```

```
PROC _____ DATA = exercice _____ ;
_____
```

```
PROC FORMAT LIBRARY = WORK;
  VALUE _____ = _____ = _____
  VALUE mois2 _____ = "Jan à Juin"
  _____ = "Juil à Déc";
RUN;
```

```
LIBNAME _____ "C:\Users\Individu\Desktop";
```

```
DATA repert _____ ;
  SET _____ ;
```

```
IF hopital _____ THEN
  region = "Capitale nationale";
ELSE _____ ;
```

```
/*imc = poids en kilogrammes divisé par la
  taille en mètres au carré.
  kg = lbs / 2.2*/
imc = _____ ;
```

```
date_naissance = _____ ;
```

```
date_conception = _____ ;
```

```
ville = _____ ;
```

```
sexe = _____ ;
```

```
IF sexe = "H" THEN _____ ;
```

```
_____ date_conception date_naissance _____  
_____ sexf. mois_nais _____ ;
```

```
DROP _____ ;
```

```
LABEL _____ ;
```

```
_____ Moyenne et compagnie _____  
PROC MEANS DATA = _____ ;
```

```
RUN ;
```

```
_____ PROC _____ DATA = _____ ;  
TABLES _____ ;
```

```
RUN ;
```

Quiz¹

PARTIE 2 – PRINCIPES DE PROGRAMMATION

2. Après l'exécution d'un code dans l'éditeur, il est important de vérifier :
 - a. L'explorateur
 - b. Les résultats
 - c. Le dossier en cours
 - d. Le journal
3. Les 3 principales composantes d'un programme SAS sont :
 - a. Code, énoncé DATA, énoncé libre
 - b. Étape DATA, étape PROC, énoncé général
 - c. Étape DATA, énoncé PROC, titre
 - d. Aucune de ces réponses
4. Dans le journal, il est important de vérifier les lignes de couleur :
 - a. Bleu et noir
 - b. Rouge et vert
 - c. Bleu et rouge
 - d. Rouge et noir
5. Dans le haut de la fenêtre active de SAS, le message « PROC step running » reste affiché longtemps, après avoir soumis la procédure PRINT. Qu'est-ce qui est arrivé ?
 - a. Une faute d'orthographe dans une option
 - b. Une option spécifiée est invalide
 - c. La procédure ne se termine pas par l'énoncé RUN;
6. Le message « WARNING : The quoted string currently being processed has become more than 262 characters long » apparaît dans le journal. Qu'est-ce qui est arrivé ?
 - a. Un nombre impair de guillemets apparaît dans un énoncé
 - b. Une faute d'orthographe dans une option
 - c. Un énoncé ne se termine pas par un point-virgule

PARTIE 3 – IMPORTATION DE DONNÉES

7. Quel nom de variable est valide ?
 - a. 1janvier
 - b. #janvier
 - c. _janvier
 - d. Janvier-01
8. Quel nom de fichier de données est un fichier permanent SAS ?
 - a. Work.donnees
 - b. Mydonnees
 - c. Temp.donnees
 - d. Aucune de ces réponses

¹ Traduit, adapté ou inspiré en partie de SAS Certification Prep Guide : Base Programming for SAS 9, 3rd edition. SAS Institute Inc. : Cary, NC. 802 pages.

PARTIE 4 – FORMATS

9. De quelle façon peut-on enregistrer des formats dans un catalogue permanent?
- LIBNAME biblio 'c:\\desktop\
PROC FORMAT LIBRARY = biblio;
 - LIBNAME biblio 'c:\\desktop\
PROC FORMAT LIBRARY;
 - LIBRARY = 'c:\\desktop\
PROC FORMAT LIBRARY;
 - LIBNAME biblio 'c:\\desktop\
PROC LIBRARY = biblio;
10. Cocher toutes les réponses qui s'appliquent : un nom de format créé dans l'énoncé VALUE doit respecter les critères suivants :
- Ne peut commencer ni terminer par un point
 - Doit être d'une longueur d'au moins 8 caractères
 - Ne peut être un format SAS
 - Doit commencer par un \$ si c'est un format pour variable alphanumérique.
 - Ne peut commencer ni terminer par un chiffre
 - Ne peut pas être le nom d'une variable SAS
11. Quel énoncé est faux? La spécification des valeurs sur lesquelles appliquer un format dans un énoncé VALUE peuvent utiliser :
- Une étendue de valeurs numériques comme 1-26
 - Une étendue de valeurs alphanumériques comme 'A'-'Z'
 - Une valeur unique comme 01 ou 'A'
 - Une liste de valeurs numériques et alphanumériques séparées par des virgules, comme 01, 'A', 2016, 'Janvier', 31

PARTIE 5 – ÉTAPE DATA

En se référant aux données ci-dessous :

Nom	Age	Sexe	Ville
John	15	H	Toronto,ON
Jane	45	F	Calgary,AB
Anne	26	F	Edmonston,NB
Sophie	.	F	Montréal,QC
Julien	31		Winnipeg,MB

12. Combien d'observations et de variables y-a-t-il?
- 5 observations, 3 variables
 - 6 observations, 4 variables
 - 5 observations, 4 variables
 - 6 observations, 3 variables
13. De quel type les variables Age et Sexe sont-elles, respectivement?
- Alphanumérique, Alphanumérique
 - Alphanumérique, Numérique
 - Numérique, Numérique
 - Numérique, Alphanumérique
 - Il est impossible de le savoir à partir des données présentes

14. Avec quel énoncé peut-on extraire la province de la variable Ville?

- a. Substr(ville,1,2);
- b. Substr(ville,10);
- c. Scan(ville,2);
- d. Scan(ville,10,2);

15. Quelle fonction utiliser pour calculer la somme des variables var1 à var4?

- a. Sum(of var1,var4)
- b. Sum(of var1-var4)
- c. Sum(var1,var4)
- d. Sum(var1-var4)

16. Quelle partie de l'énoncé suivant est évaluée en premier?

IF age > 30 AND (sexe = "F" or (sexe = "M" and ville = "Québec")) THEN groupe = 1;

- a. IF age>30
- b. Sexe = "F"
- c. sexe = "M" and ville = "Québec"
- d. THEN groupe = 1;

17. Quel est l'avantage que les dates soient enregistrées sous forme de variables numériques (nombre de jours depuis le 1^{er} janvier 1960)?

- a. Peuvent être utilisées avec des fonctions pour manipuler des caractères
- b. Peuvent être utilisées avec des opérateurs mathématiques
- c. Plus facile à corriger
- d. Plus facile à lire et interpréter

18. Identifier le ou les énoncé(s) effectuant des opérations équivalentes aux énoncés suivants :

If province = "QC" THEN langue = "Français";
IF province = "NB" THEN langue = "Français";
IF province ^= "QC" and province ^= "NB" THEN langue = "Anglais";

- a. IF province in ("QC","NB") THEN langue = "Français";
ELSE langue = "Anglais";
- b. IF province in ("QC","NB") THEN langue = "Français";
ELSE IF province NOT IN {"QC","NB"} THEN langue = "Anglais";
- c. If province = "QC" THEN langue = "Français";
IF province = "NB" THEN langue = "Français";
ELSE IF province ^= "NB" THEN langue = "Anglais";
- d. IF province ^= "qc" and province ^= "nb" THEN langue = "Anglais";
ELSE langue = "Français";

19. Quelles variables feront partie des données créées à l'aide du code suivant?

```
DATA work.nettes (DROP = id adresse);  
SET biblio.brutes (KEEP = id sexe age imc atcd adresse);  
IF 18<=age<=60;  
RUN;
```

- a. Id, sexe, age, imc, atcd, adresse
- b. Id, sexe, age, imc, atcd
- c. Sexe, age, imc, atcd
- d. Sexe, imc, atcd

20. Qu'arrivera-t-il en juxtaposant les deux jeux de données suivants?

```
DATA toto;  
  MERGE donnees1 donnees2;  
  BY id;  
RUN;
```

Donnees1		
Id	Age	Sexe
1	28	F
7	48	F
5	36	H

Donnees2		
Id	Age	IMC
1	28	24
4	42	25
7	48	23

- Des erreurs seront affichées dans le journal
- Le fichier toto comprendra 2 observations
- Le fichier toto comprendra 4 observations
- Il y aura des données manquantes dans toto

PARTIE 6 – PROCÉDURE

21. On peut placer l'énoncé FORMAT dans une étape DATA ou PROC. Qu'est-ce qui arrive lorsqu'on le place dans une étape PROC?

- Des formats sont associés temporairement aux variables, si elles n'en avaient pas déjà un.
- Des formats sont associés de façon permanente aux variables
- Les formats deviennent accessibles à d'autres jeux de données
- Des formats sont associés temporairement aux variables, en écrasant leur format préséparé si elles en avaient déjà un

22. L'énoncé « BY age DESCENDING IMC sexe » de la procédure SORT triera :

- D'abord en ordre décroissant selon l'âge, puis en ordre croissant selon l'IMC et le sexe
- D'abord en ordre croissant selon l'âge, puis en ordre décroissant selon l'IMC et finalement en ordre croissant selon le sexe
- D'abord en ordre croissant selon l'âge, puis en ordre décroissant selon l'IMC et le sexe

Solutions :

Exercice 1

```
/*!!! suppose que le fichier Excel se trouve dans le dossier en cours*/
PROC IMPORT DATAFILE = ".\Atelier SAS Données.xlsx"
  OUT = exercice
  DBMS = excel
  REPLACE;
  SHEET = "Exercice$A1:J11";
RUN;

PROC CONTENTS DATA = exercice VARNUM;
RUN;

PROC FORMAT LIBRARY = WORK;
  VALUE $sexf "F" = "Fille" "G" = "Garçon";
  VALUE mois2f low-6 = "Jan à Juin" 7-12 = "Juil à Déc";
RUN;

LIBNAME repert "C:\Users\AnJu001\Desktop";

DATA repert.nouvelle;
  SET exercice;

  IF hopital in ("CHUL","HDL","HSFA") THEN region = "Capitale nationale";
  ELSE region = "Ailleurs";

  /*imc = poids en kg divisé par la taille en m au carré. kg = lbs / 2.2*/
  imc = (poids/2.2) / (grandeur/100) **2;

  date_naissance = MDY(mois_nais,jour_nais,an_nais);
  date_conception = date_naissance - age_gest_jrs;

  ville = SCAN(coord_mere,3,"");

  sexe = SUBSTR(NAM,11,1);
  IF sexe = "H" THEN sexe = "G";

  FORMAT date_conception date_naissance mmddyy8. sexe sexf. mois_nais mois2f.;

  DROP an_nais jour_nais nam coord_mere ;
  LABEL age_gest_jrs = "Âge gestationnel (jrs)";
RUN;

TITLE1 "Moyenne et compagnie";
PROC MEANS DATA = repert.nouvelle MAXDEC = 3;
  VAR imc poids grandeur age_gest_jrs;
RUN;

TITLE1 "Stats descriptives";
TITLE2 "Fréquences";
PROC FREQ DATA = repert.nouvelle;
  TABLES ville sexe*mois_nais;
RUN;
```

Quiz

2.d 3.b 4.b 5.c 6.a 7.c 8.c 9.a 10.acde 11.d 12.c 13.d 14.c 15.b 16.c 17.b 18.ab 19.c
20.a 21.d 22.b