

# STT1682 – Progiciels en Statistique et Actuariat

## Examen Intra – 13 Mars 2013

Chargé de cours : David Langlois  
Documentation papier est permise  
Durée de l'examen : 2h  
Pondération : 30%

### **#1 Questions à choix multiples (20 pts) :**

**Note :** Une ou plusieurs lettres peuvent être sélectionnés par question

**1.1)** Lesquels de ces noms de variables sont des noms valides en SAS :

- a) 50Prime
- b) Prime50
- c) Prime\_Annuelle
- d) PrimeAnnuelleChargeeAuQuebec
- e) Prime%

**1.2)** Par défaut, lesquels de ces procédures affichent des résultats dans le fenêtre Sortie :

- a) PROC SUMMARY
- b) PROC FREQ
- c) PROC RANK
- d) PROC SORT
- e) PROC CONTENTS

**1.3)** Quelle est la date SAS représentée par le chiffre 15 :

- a) 16 Janvier, 0
- b) 16 Janvier, 2000
- c) 16 Janvier, 1900
- d) 16 Janvier, 1960
- e) 16 Janvier, 1000

**1.4)** Si on sommarise une base de données à l'aide d'un PROC SUMMARY sans l'option NWAY avec 3 variables dans la déclaration CLASS et chacune de ces variables sont composées de 10 valeurs différentes et que toute combinaison est possible, combien d'observations aura la base résultante :

- a) 1,331 (1000+300+30+1)
- b) 1,300
- c) 1,031
- d) 1,000
- e) 300

## **#1 Questions à choix multiples (Suite) :**

**1.5)** Pour une base de donnée à 1000 observations, combien de fois sera exécuté l'opération  $X=X+1$ ; par l'ordinateur dans le code ci-dessous :

```
...  
DO I=1 TO 1000 BY 2;  
X=X+1;  
END;  
...
```

- a) 1,000
- b) 500
- c) 1,500
- d) 1,000,000
- e) 500,000 (500\*1000)

**1.6)** Une déclaration ARRAY créant un vecteur à 10 positions peut être remplacé par une déclaration LENGTH créant 10 variables (**Vrai/Faux**)

Vrai, une déclaration ARRAY ne fait que créer plusieurs variables de longueur X, on pourrait la remplacer par une déclaration LENGTH.

**1.7)** À l'intérieur d'une étape DATA, il est possible d'effectuer une opération SAS utilisant l'information de n'importe quelle observation (**Vrai/Faux**)

Faux, seulement l'information de l'observation courante peut être utilisée à l'intérieure d'une étape DATA.

**1.8)** Il n'y a aucune différence à utiliser l'option WHERE avec la déclaration DATA ou avec la déclaration SET (**Vrai/Faux**)

Faux, l'option WHERE à l'étape SET sera exécutée à la lecture de la base de donnée vs à l'écriture de la base de donnée.

**1.9)** On peut lire plusieurs bases de données avec une seule étape DATA (**Vrai/Faux**)

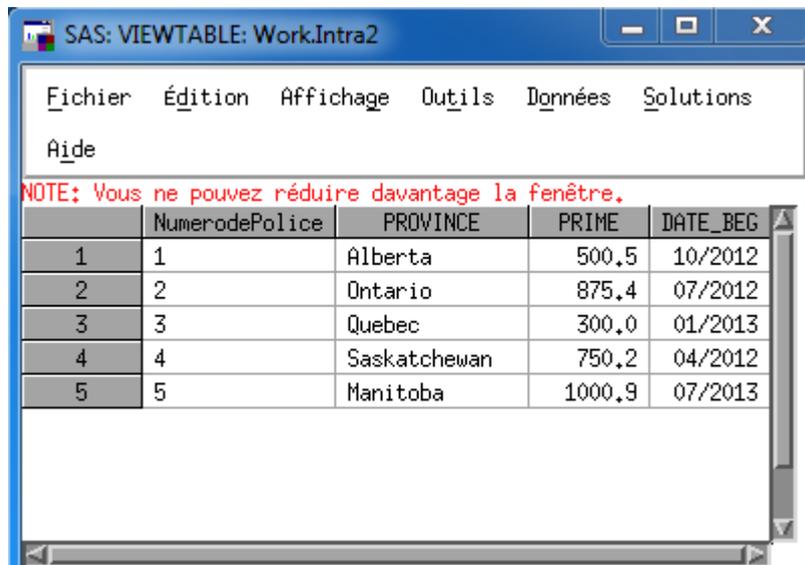
Vrai

**1.10)** Pour enlever les doublons d'une base de donnée, il est préférable d'utiliser un PROC SUMMARY qu'un PROC SORT (**Vrai/Faux**)

Vrai, un PROC SUMMARY sera beaucoup plus rapide et efficace qu'un PROC SORT pour ce genre de traitement.

## #2 Base de Donnée SAS (30 pts) :

a) Écrire le code nécessaire pour créer la base de données ci-dessous avec exactement les mêmes variables, valeurs et propriétés :



	NumerodePolice	PROVINCE	PRIME	DATE_BEG
1	1	Alberta	500,5	10/2012
2	2	Ontario	875,4	07/2012
3	3	Quebec	300,0	01/2013
4	4	Saskatchewan	750,2	04/2012
5	5	Manitoba	1000,9	07/2013

Voici le résultat d'un PROC CONTENT sur cette base de données :

### Liste alphabétique des variables et des attributs

N	Variable	Type	Long.	Format	Libellé
4	DATE_BEG	Num.	8	MMYY5.	
1	POLICE	Texte	8		NumerodePolice
3	PRIME	Num.	8	8.1	
2	PROVINCE	Texte	12		

### Informations sur le tri

Trié par POLICE  
Validé OUI  
Jeu de caractères ASCII

Solution : (les journées ne sont pas importantes pour DATE\_BEG)

Note : Plusieurs solutions seront acceptés

```
DATA Intra2;
INPUT POLICE $8. PROVINCE $12. PRIME 8.DATE_BEG MMDDYY10.;
FORMAT PRIME 8.1 DATE_BEG MMYYS.;
DATALINES;
1 Alberta 500.5 10/01/2012
2 Ontario 875.4 07/01/2012
3 Quebec 300.0 01/01/2013
4 Saskatchewan 750.2 04/01/2012
5 Manitoba 1000.9 07/01/2013
;
RUN;
```

```
PROC SORT DATA=Intra2;
BY POLICE;
RUN;
```

**b)** Si la valeur de DATE\_BEG pour la 1ère observation est égale à 19281 et que le chiffre correspondant au 1er Novembre 2012 est 19298 , écrire sous le format DDMMYYD10. la 1ère observation de DATE\_BEG

Solution : 15-10-2012

**c)** Écrire la base résultante si le WHERE ci-dessous est appliqué sur la base ci-dessus :

```
WHERE=((LENGTH(PROVINCE) EQ 7 OR FIND(PROVINCE,'o') NOT EQ 0)
AND MONTH(DATE_BEG) LT 10)
```

Solution :

2	Ontario	875.4	07/2012
5	Manitoba	1000.9	07/2013

**d)** Quel est le résultat de l'opération suivante pour chacune des 5 observations :

```
SUBSTR(PUT(DATE_BEG,DDMMYY10.),4,3) || SUBSTR(PUT(DATE_BEG,DDMMYY10.),9,2)
DATE_BEG sous format DDMMYY10.= "JJ/MM/AAAA"
SUBSTR(PUT(DATE_BEG,DDMMYY10.),4,3) = "MM/"
SUBSTR(PUT(DATE_BEG,DDMMYY10.),9,2) = "AA" (2 derniers digits)
```

Solution :

10/12  
07/12  
01/13  
04/12  
07/13

### **#3 Efficacité de Programmation (20 pts) :**

Écrire un programme obtenant le même résultat final que le code ci-dessous (la base PRIME6), mais de la façon la plus efficace possible :

```
DATA PRIME2;
SET PRIME1;

IF PRIME GT 0;

IF YEAR(DATE_EFF) EQ 2012;

PRIME1=1/12*PRIME_TOT;PRIME2=1/12*PRIME_TOT;
PRIME3=1/12*PRIME_TOT;PRIME4=1/12*PRIME_TOT;
PRIME5=1/12*PRIME_TOT;PRIME6=1/12*PRIME_TOT;
PRIME7=1/12*PRIME_TOT;PRIME8=1/12*PRIME_TOT;
PRIME9=1/12*PRIME_TOT;PRIME10=1/12*PRIME_TOT;
PRIME11=1/12*PRIME_TOT;PRIME12=1/12*PRIME_TOT;

RUN;

PROC SORT DATA=PRIME2 OUT=PRIME3;
BY PROVINCE;
RUN;

PROC SUMMARY DATA=PRIME3 NWAY MISSING;
CLASS PROVINCE;
VAR PRIME1 PRIME2 PRIME 3 PRIME4 PRIME5 PRIME6
PRIME7 PRIME8 PRIME9 PRIME10 PRIME11 PRIME12;
OUTPUT OUT=PRIME4 SUM=;
RUN;

PROC SUMMARY DATA=PRIME3 NWAY MISSING;
CLASS PROVINCE;
VAR PRIME_TOT;
OUTPUT OUT=PRIME5 MEAN=;
RUN;

DATA PRIME6;*Resultat Final;
SET PRIME4;
SET PRIME5;
IF PRIME_TOT GE 1000;
DROP _TYPE_ _FREQ_;
RUN;
```

### Solution #3

Note : Plusieurs solutions seront acceptés

```
DATA PRIME2;
SET PRIME1(WHERE=PRIME GT 10 AND YEAR(DATE_EFF)=2012);

ARRAY PRIME{12} 8.0 PRIME1-PRIME12;
DO I=1 TO 12;
PRIME{ I}=1/12*PRIME_TOT;
END;

RUN;

PROC SUMMARY DATA=PRIME2 NWAY MISSING;
CLASS PROVINCE;
VAR PRIME1-PRIME12 PRIME_TOT;
OUTPUT OUT=PRIME3(DROP=_TYPE__FREQ_ WHERE=(PRIME_TOT GE 1000))
SUM(PRIME1-PRIME12)= MEAN(PRIME_TOT)=;
RUN;
```

### **#4 Trouver et corriger les erreurs dans le code ci-dessous (10 pts):**

(encercler les erreurs corrigées dans votre cahier réponse)

Note : Il peut exister plusieurs corrections possibles aux erreurs.

```
DATA Intra4_2;
SET Intra4_1 (WHERE=( _N_ => 1000 ) );
BY PROVINCE;
IF FIRST.PROVINCE;
IF PROVINCE="Quebec" THEN DO;
PRIME= PRIME**0.5;
RABAIS="OUI";
PROV2=SUBSTR(PROVINCE,1,3);
END;
ELSE DO; PRIME=PRIME*2; END;
RENAME PROVINCE=PROV;
RUN;
```

### **#5 Programmation Avancé (20 pts) :**

À partir de la base de données **Intra5** décrite ci-dessous, écrire le code nécessaire

pour créer une base de données avec les caractéristiques suivantes:

- 1 observation pour chaque valeur différente de VILLE et seulement deux variables (VILLE et COEFF\_VAR)
- Observations où DATE\_COMPTABLE est strictement inférieure à 2009 doivent être exclues
- Créer la variable PRIME2 qui est égale à PRIME multiplié par un facteur variant par année de DATE\_COMPTABLE :

<b>Année</b>	<b>Facteur</b>
2009	1.05
2010	1.08
2011	1.10
2012	1.15

- Une variable COEFF\_VAR représentant le coefficient de variation (écart type divisé par moyenne) de la variable PRIME2 pour chaque valeur de VILLE
- COEFF\_VAR doit être trié de façon décroissante pour chaque valeur de VILLE

### **Base Intra5**

#### Variable 1

Nom : VILLE  
Length : \$10.

#### Variable 2

Nom : DATE\_COMPTABLE  
Length : \$8.  
Valeurs : Chaîne de caractères de longueur 8 représentant la date sous forme : "AAAAMMJJ"

#### Variable 3

Nom : PRIME  
Length : 8.0

### **Solution #5**

Note : Plusieurs solutions seront acceptés

```

DATA Intra_5_Step1;
SET Intra5(WHERE=(INPUT(SUBSTR(DATE_COMPTABLE,1,4),8.0) GE 2009));

IF SUBSTR(DATE_COMPTABLE,1,4)=2009 THEN PRIME2=PRIME*1.05;
ELSE IF SUBSTR(DATE_COMPTABLE,1,4)=2010 THEN PRIME2=PRIME*1.08;
ELSE IF SUBSTR(DATE_COMPTABLE,1,4)=2011 THEN PRIME2=PRIME*1.10;
ELSE IF SUBSTR(DATE_COMPTABLE,1,4)=2012 THEN PRIME2=PRIME*1.15;

RUN;

PROC SUMMARY DATA=Intra_5_Step1 NWAY MISSING;
CLASS VILLE;
VAR PRIME2;
OUTPUT OUT= Intra_5_Step2(DROP=_TYPE__FREQ_)
STD(PRIME2)=STD_PRIME2 MEAN(PRIME2)=MEAN_PRIME2;
RUN;

DATA Intra_5_Step3(DROP=STD_PRIME2 MEAN_PRIME2);
SET Intra_5_Step2;
COEFF_VAR= STD_PRIME2/MEAN_PRIME2;
RUN;

PROC SORT DATA= Intra_5_Step3;
BY DESCENDING COEFF_VAR; *BY VILLE DESCENDING COEFF_VAR aussi accepte;
RUN;

```