



# APPRENTISSAGE DU LANGAGE DE PROGRAMMATION SAS

Gramondo Martina

UT1 FOAD M2 & DU  
2018/19





**SEMAINE 3:**  
AUTOFORMATION PROCÉDURES CONTRÔLE QUALITÉ DES DONNÉES  
PROC FREQ

Gramondo Martina

UT1 FOAD M2 & DU  
2018/19



# PROCÉDURES CONTRÔLE QUALITÉ

Une tâche fondamentale dans un projet est garantir la qualité des données (data management, data quality).

Pour contrôler les données on utilisera:

- Des logical check, des contrôle croisés entre les variables pour valider des valeurs des observations, les experts métier mettront en place toute une liste des contrôles à faire pour détecter des anomalies
- Des étapes des PROC pour contrôler les valeurs extrêmes et les modalités des variables (problèmes dans la casse, valeurs pas acceptables, problèmes d'orthographe des valeurs,...)

Pour la partie PROC on utilisera:

- Sur les variables numériques:
  - PROC MEANS
  - PROC UNIVARIATE
- Sur les variables caractères et numériques discrètes:
  - PROC FREQ

Ces procédures sont comprises dans le module SAS/BASE que vous avez, bien évidemment.

# PROC FREQ

La PROC FREQ est une procédure très complète ici l'help [PROC FREQ DOC](#)

# EXERCICES

## SEMAINE 3

- Lire la doc de la PROC FREQ
- Déclarer la bibliothèque UT1\_FOAD (les tables des exercices)
- `libname UT1_SAS '/folders/myfolders/UT1_FOAD' ;`
- A partir de la table UT1\_SAS.CUSTOMER\_DIM utilisez la PROC FREQ pour générer la sortie suivante sur les variables Customer\_Country et Customer\_Gender



**KEEP  
CALM  
AND  
DO YOUR  
HOMEWORK**

Nombre de niveaux de variables		
Variable	Libellé	Niveaux
Customer_Country	Customer Country	7
Customer_Gender	Customer Gender	2

Customer Country		
Customer_Country	Fréquence	Pourcentage
US	28	36.36
CA	15	19.48
DE	10	12.99
AU	8	10.39
TR	7	9.09
IL	5	6.49
ZA	4	5.19

Customer Gender		
Customer_Gender	Fréquence	Pourcentage
M	47	61.04
F	30	38.96

# EXERCICES

## SEMAINE 3



**KEEP  
CALM  
AND  
DO YOUR  
HOMEWORK**

- A partir de la table UT1\_SAS.CUSTOMER\_DIM utilisez la PROC FREQ pour générer la sortie suivante sur les variables Customer\_Gender et Customer\_Age\_Group

Table de Customer_Gender par Customer_Age_Group					
Customer_Gender	Customer_Age_Group	Fréquence	Pourcentage	Pourcent. ligne	Pourcent. colonne
F	15-30 years	12	15.58	40.00	54.55
	31-45 years	12	15.58	40.00	44.44
	46-60 years	4	5.19	13.33	28.57
	61-75 years	2	2.60	6.67	14.29
	Total	30	38.96	100.00	
M	15-30 years	10	12.99	21.28	45.45
	31-45 years	15	19.48	31.91	55.56
	46-60 years	10	12.99	21.28	71.43
	61-75 years	12	15.58	25.53	85.71
	Total	47	61.04	100.00	
Total	15-30 years	22	28.57		100.00
	31-45 years	27	35.06		100.00
	46-60 years	14	18.18		100.00
	61-75 years	14	18.18		100.00
	Total	77	100.00		

# EXERCICES

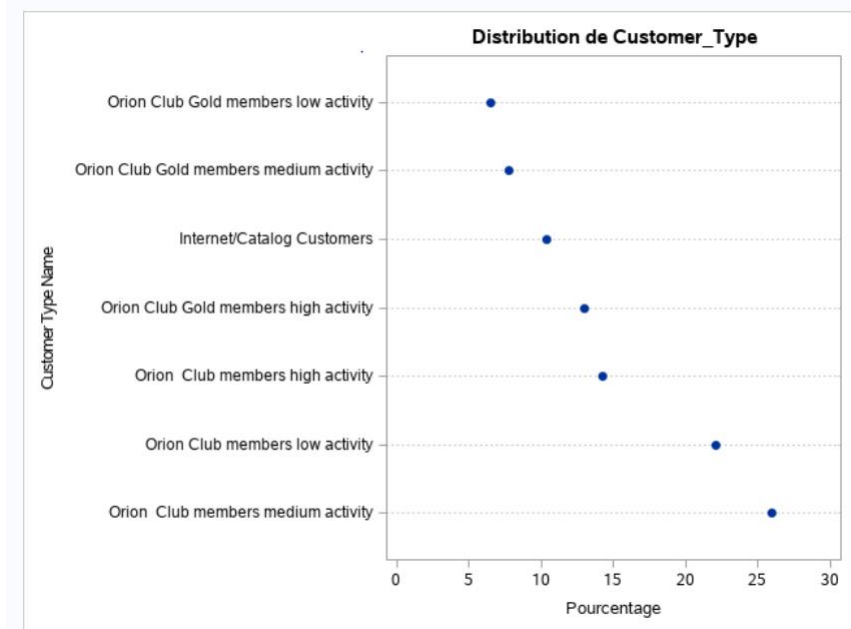
## SEMAINE 3



KEEP  
CALM  
AND  
DO YOUR  
HOMEWORK

- A partir de la table UT1\_SAS.CUSTOMER\_DIM utilisez la PROC FREQ pour générer la sortie suivante sur la variable Customer\_type

Customer Type Name				
Customer_Type	Fréquence	Pourcentage	Fréquence cumulée	Pourcentage cumulé
Orion Club members medium activity	20	25.97	20	25.97
Orion Club members low activity	17	22.08	37	48.05
Orion Club members high activity	11	14.29	48	62.34
Orion Club Gold members high activity	10	12.99	58	75.32
Internet/Catalog Customers	8	10.39	66	85.71
Orion Club Gold members medium activity	6	7.79	72	93.51
Orion Club Gold members low activity	5	6.49	77	100.00



# EXERCICES

## SEMAINE 3



**KEEP  
CALM  
AND  
DO YOUR  
HOMEWORK**

- BONUS: A partir de la table UT1\_SAS.CUSTOMER\_DIM utilisez la PROC FREQ et les ODS TRACE et ODS SELECT pour générer seulement la sortie suivante:

Nombre de niveaux de variables		
Variable	Libellé	Niveaux
Customer_Country	Customer Country	7
Customer_Gender	Customer Gender	2
Customer_Age_Group	Customer Age Group	4
Customer_Type	Customer Type Name	7
Customer_Group	Customer Group Name	3





## REPONSES DANS LE FORUM

Proposées vos réponses dans le forum afin de pouvoir en parler avec les autres élèves

Vos réponses ne seront pas notées, votre participation au forum oui

Posez des questions dans le forum

Lien pour l'help SAS: