



Chapitre 4: Les fonctions et les procédures statistiques

SAS- M1 Actuariat / ES

Nassira CHEKKAI

nassira.chekkai@univ-lyon1.fr

**Université de Lyon 1 –
Institut de Science Financière et d'Assurances
ISFA**

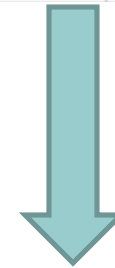
Novembre 2022

1. Opérations sur les variables

A. Renommer les variables (l'instruction RENAME)

```
*Importation store.csv*/
PROC IMPORT OUT= WORK.store
              DATAFILE=
"C:\MySAS\store.csv"
              DBMS=CSV
REPLACE;
      GETNAMES=yes;
      DATAROW=2;
RUN;
PROC PRINT;
RUN;
/*Renommer des variables*/
DATA store_v2;
SET store;
RENAME storeID = ID
        storetypeID=storetype;
RUN;
PROC PRINT;
RUN;
```

Obs.	storeID	storetypeID	storeSize	cityID
1	S0091	ST04	19	C013
2	S0012	ST04	28	C005
3	S0045	ST04	17	C008
4	S0032	ST03	14	C019



Obs.	ID	storetype	storeSize	cityID
1	S0091	ST04	19	C013
2	S0012	ST04	28	C005
3	S0045	ST04	17	C008
4	S0032	ST03	14	C019
5	S0027	ST04	24	C022

1. Opérations sur les variables

B. Sélectionner les variables à afficher dans la nouvelle base de données (l'instruction KEEP)

```
/*Sélectionner les variables à afficher  
dans la nouvelle base de données*/  
DATA store_V3;  
SET store;  
KEEP storeID;  
RUN;  
PROC PRINT;  
RUN;
```

Obs.	storeID
1	S0091
2	S0012
3	S0045
4	S0032
5	S0027

2. Fonctions statistiques

- La fonction **min** retourne le nombre minimum.
- La fonction **max** retourne le nombre maximum.
- La fonction **mean** retourne la moyenne.
- La fonction **sum** retourne la somme.
- La fonction **std** retourne l'écart type.
- La fonction **median** retourne la médiane.

2. Fonctions statistiques

```
/*fonctions statistiques*/
```

```
DATA notes;
```

```
INPUT nom$ note1 note2 note3;
```

```
CARDS;
```

```
SARA 13 14 16
```

```
PAUL 17 10 9
```

```
VICTOR 14 15 12
```

```
;
```

```
RUN;
```

```
DATA note_V2;
```

```
SET notes;
```

```
Max=max(of note1-note3);
```

```
Min=min(of note1-note3);
```

```
Moyenne=mean(of note1-note3);
```

```
Somme=sum(of note1-note3);
```

```
Standard_Deviation=std(of note1-note3);
```

```
Median=median(of note1-note3);
```

```
RUN;
```

```
PROC PRINT;
```

```
RUN;
```

Obs.	nom	note1	note2	note3	Max	Min	Moyenne	Somme	Standard_Deviation	Median
1	SARA	13	14	16	16	13	14.3333	43	1.52753	14
2	PAUL	17	10	9	17	9	12.0000	36	4.35890	10
3	VICTOR	14	15	12	15	12	13.6667	41	1.52753	14

2. Fonctions statistiques

- **CDF** Renvoie une valeur à partir d'une distribution de probabilité cumulée.
 - **PDF** Renvoie une valeur à partir d'une distribution de densité de probabilité (masse).
 - **QUANTILE** Renvoie le quantile d'une distribution lorsque vous spécifiez la probabilité de gauche (CDF).
-
- Liens vers d'autres fonctions SAS:
<https://support.sas.com/documentation/cdl/en/lrdict/64316/HTML/default/viewer.htm#a000270634.htm>

2. Fonctions statistiques

```
/* Fonctions Statistique 2
*/
DATA NormalDistribution;
cdf=CDF('NORMAL',0,0,1);
pdf=PDF('NORMAL',0,0,1);
quantile=QUANTILE('NORMAL',0
.5,0,1);
RandomNumber=RAND('NORMAL',0
,1);
RUN;
PROC PRINT;
RUN;
```

Obs.	cdf	pdf	quantile	RandomNumber
1	0.5	0.39894	0	-0.14985

3. Procédure FREQ

- La procédure FREQ affiche les valeurs d'une variable catégorielle donnée dans la fenêtre Sortie, ainsi que les nombres et les proportions.
- La procédure FREQ peut fonctionner avec des variables qualitatives de type chaîne (caractère) ou numériques.

3. Procédure FREQ

➤ Affichage des fréquences des valeurs

```

/*Procédure FREQ*/
DATA notes;
INPUT nom$ note1 note2 note3 mention$;
CARDS;
SARA 13 14 16 "admis"
PAUL 17 10 9 "admis"
VICTOR 4 5 7 "ajournee"
;
RUN;
/*Table de fréquence*/
PROC FREQ DATA = notes ;
RUN;
PROC PRINT;
RUN;

```

Obs.	nom	note1	note2	note3	mention
1	SARA	13	14	16	"admis"
2	PAUL	17	10	9	"admis"
3	VICTOR	4	5	7	"ajourne"

nom	Fréquence	Pourcentage	Fréquence cumulée	Pourcentage cumulé
PAUL	1	33.33	1	33.33
SARA	1	33.33	2	66.67
VICTOR	1	33.33	3	100.00

note1	Fréquence	Pourcentage	Fréquence cumulée	Pourcentage cumulé
4	1	33.33	1	33.33
13	1	33.33	2	66.67
17	1	33.33	3	100.00

note2	Fréquence	Pourcentage	Fréquence cumulée	Pourcentage cumulé
5	1	33.33	1	33.33
10	1	33.33	2	66.67
14	1	33.33	3	100.00

note3	Fréquence	Pourcentage	Fréquence cumulée	Pourcentage cumulé
7	1	33.33	1	33.33
9	1	33.33	2	66.67
16	1	33.33	3	100.00

mention	Fréquence	Pourcentage	Fréquence cumulée	Pourcentage cumulé
"admis"	2	66.67	2	66.67
"ajourne"	1	33.33	3	100.00

3. Procédure FREQ

➤ Affichage des tables de fréquence des variables sélectionnées

```
/*Table de fréquence des variables  
sélectionnées*/  
PROC FREQ DATA = notes ;  
TABLES mention;  
RUN;  
PROC PRINT;  
RUN;
```

mention	Fréquence	Pourcentage	Fréquence cumulée	Pourcentage cumulé
"admis"	2	66.67	2	66.67
"ajourne"	1	33.33	3	100.00

3. Procédure FREQ

➤ **Affichage de la table de contingence (étudier l'interaction entre deux variables catégorielles).**

```
/*Table de contingence */
PROC FREQ DATA = notes ;
TABLES note1*mention;
RUN;
PROC PRINT;
RUN;
```

Obs.	nom	note1	note2	note3	mention
1	SARA	13	14	16	"admis"
2	PAUL	17	10	9	"admis"
3	VICTOR	4	5	7	"ajourne"

La procédure FREQ

Fréquence Pourcentage Pct de ligne Pct de col.	Table de note1 par mention		
	mention		
	note1	"admis"	"ajourne"
4	0	1	1
	0.00	33.33	33.33
	0.00	100.00	
	0.00	100.00	
13	1	0	1
	33.33	0.00	33.33
	100.00	0.00	
	50.00	0.00	
17	1	0	1
	33.33	0.00	33.33
	100.00	0.00	
	50.00	0.00	
Total	2	1	3
	66.67	33.33	100.00

4. Procédure MEANS

La procédure MEANS permet d'afficher des statistiques descriptives sur les variables (Moyenne, Écart-type, Minimum et Maximum).

```
/*Statistiques*/  
PROC MEANS DATA =  
notes ;  
RUN;  
PROC PRINT;  
RUN;
```

La procédure MEANS

Variable	N	Moyenne	Ec-type	Minimum	Maximum
note1	3	11.3333333	6.6583281	4.0000000	17.0000000
note2	3	9.6666667	4.5092498	5.0000000	14.0000000
note3	3	10.6666667	4.7258156	7.0000000	16.0000000

4. Procédure MEANS

➤ Statistiques en sélectionnant les variables

```
/*Statistiques en  
sélectionnant les variables*/  
PROC MEANS DATA = notes ;  
VAR note1;  
RUN;  
PROC PRINT;  
RUN;
```

La procédure MEANS

Variable d'analyse : note1				
N	Moyenne	Ec-type	Minimum	Maximum
3	11.3333333	6.6583281	4.0000000	17.0000000

Bibliographie

- Pierre-Olivier Goard, Chap 2: Data Manipulation, Université de Lyon 1 - Institut de Science Financière et d'Assurance, M1 Actuariat/ES, 2019.
- SAS TUTORIALS: FREQUENCY TABLES USING PROC FREQ, <https://libguides.library.kent.edu/SAS/Frequencies> (Dernier accès novembre 2022).