



Initiation à SAS

Plan

- ▶ Présentation
- ▶ Présentation des fenêtres
- ▶ L'étape « DATA »
 - Lire/Ecrire un fichier, formats
 - Modifier les variables, créer de nouvelles variables, fonctions mathématiques
 - Les instructions de base (if, while, drop, retain, ...)
 - Gestion de plusieurs fichiers (set, merge)
- ▶ L'étape « PROC »
 - Les procédures de base
 - Afficher les données : print
 - Trier les données : sort
 - Décrire les données : means, freq, univariate
 - Créer, Transformer les données : standard
 - Procédures graphiques : gplot, gchart
 - Les procédures d'analyse statistiques
 - Analyse de variance : glm, mixed, npar1way
 - Régression : reg, nlin
 - Corrélation : corr

SAS



Jim Barr



Jim Goodnight



John Sall



Jane Helwig

- ▶ Date de création : 1976
- ▶ CA monde 2010 : 2,43 milliards de dollars
 - + 11 000 employés
 - + 400 agences SAS dans le monde
 - + 50 000 sites clients répartis dans 127 pays
 - 24% du CA réinvesti en R&D en 2010

Les 10 commandements



fenêtre à la française



fenêtre à jalousies



fenêtre basculante



Les différentes fenêtres

Exécution du programme

Méfiance sur les lignes rouges et vertes

Résultats



```

SAS: Log-Untitled
File Edit View Tools Solutions Help
140 L=B*K;
141 TI=(1/K)*log(B);
142 A=p0c*exp(B);

NOTE: There were 4828 observations read from the data set WORK.SORT
NOTE: The data set WORK.SORTIE4 has 4 observations and 82 variables
NOTE: DATA statement used:
      real time      0.31 seconds
      cpu time       0.01 seconds

143 proc print; var souche sexe L TI A;
144
145 run;

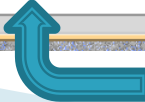
NOTE: There were 4 observations read from the data set WORK.SORTIE4
NOTE: PROCEDURE PRINT used:
      real time      1.75 seconds
      cpu time       0.01 seconds

```

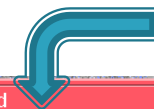
```

SAS: Program Editor-Cameroun.sas
File Edit View Tools Run Solutions Help
Autosave complete
00001 data a;
00002 infile "Cameroun-tarseemplume.csv" firstobs=3 dlm=';';
00003 input p0m p0f p1m p1f p2m p2f p3m p3f p4m p4f p5m p5f p6m p6f
00004 p7m p7f p8m p8f p9m p9f p10m p10f p11m p11f p12m p12f
00005 p13m p13f p14m p14f p15m p15f p16m p16f ;
00006 truc=1;
00007 souche='TarseEmplume';
00008 proc sort; by truc;
00009 proc means noprint;
00010 output out=meansa VAR=
00011 m0m m0f m1m m1f m2m m2f m3m m3f m4m m4f m5m m5f m6m m6f
00012 m7m m7f m8m m8f m9m m9f m10m m10f m11m m11f m12m m12f

```



Programme



```

SAS: Output-Untitled
File Edit View Tools Solutions Help

The SAS System

Obs   souche      sexe    L
-----
1     CouNu        2       0.062137
2     Huppee       2       0.062738
3     Normale      2       0.062382
4     TarseEmplume 2       0.062764

```

fenêtre à la française



fenêtre à jalousies



fenêtre basculante



Les différentes fenêtres

Erreur dans le nom du fichier

SAS est fâché !



```

SAS: Log-Untitled
File Edit View Tools Solutions Help
real time          1.61 seconds
cpu time           0.26 seconds

1  data a;
2  infile "Cwmeroun-tarseemplume.csv" firstobs=3 dlm=';';
3  input p0m p0f p1m p1f p2m p2f p3m p3f p4m p4f p5m p5f p6m p6f
4     p7m p7f p8m p8f p9m p9f p10m p10f p11m p11f p12m p12f
5     p13m p13f p14m p14f p15m p15f p16m p16f ;
6  truc=1;
7  souche='TarseEmplume';
8  run;

ERROR: Physical file does not exist, /home1/sra/users/sgrasteau/Cwm
NOTE: The SAS System stopped processing this step because of errors
WARNING: The data set WORK.A may be incomplete.  When this step was
        variables.
NOTE: DATA statement used:
real time          0.43 seconds
cpu time           0.01 seconds

```

fenêtre à la française



fenêtre à jalousies



fenêtre basculante



Les différentes fenêtres

Erreur dans le nom l'option

SAS peut aussi se montrer compréhensif



```

SAS: Program Editor-seveso.sas
File Edit View Tools Run Solutions Help
NOTE: 7 Lines submitted.
00030
00031 data seveso;
00032 infile "FormationSAS/DonneesSEVESO.csv" firstobs=2 dlm=';';
00033 input departement :$23. autorisation sevesoHR sevesoRM region :$13.
00034      pop2009 superficie taille :$10. densite densiteb :$19.;
00035 proc corr spearman;
00036 var autorisation sevesoHR sevesoRM pop2009 superficie densite ;
00037 run;
00038

SAS: Log-Untitled
File Edit View Tools Solutions Help

real time          0.21 seconds
cpu time           0.02 seconds

5   proc corr spearman;
6   var autorisation sevesoHR sevesoRM pop2009 superficie densite ;
7   run;

WARNING 1-322: Assuming the symbol SPEARMAN was misspelled as spearman.

```

fenêtre à la française



fenêtre à jalousies



fenêtre basculante



Les différentes fenêtres



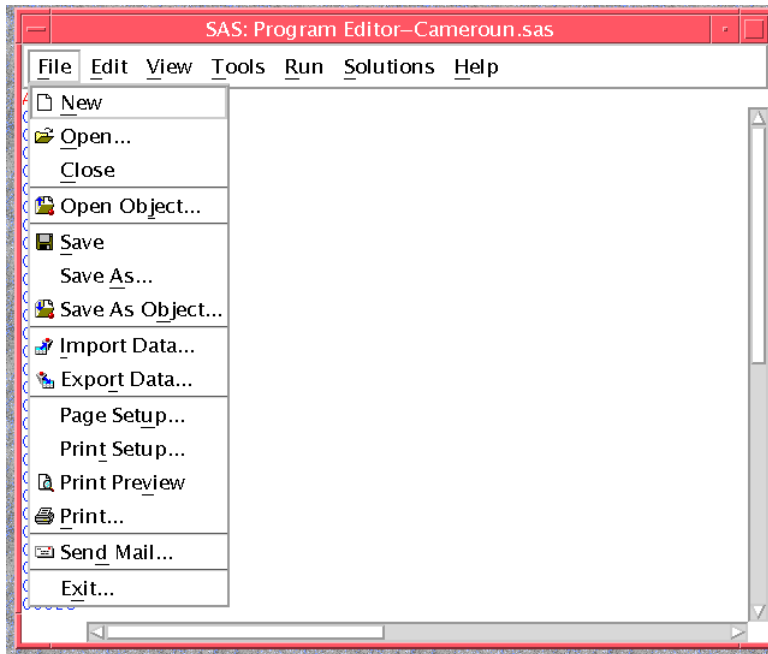
Contient les fichiers temporaires créés par SAS pendant la session



Principales fonctions du menu des autres fenêtres



Le menu



- New : créer un nouveau fichier
- Open : ouvrir un programme existant
- Close : Ferme la fenêtre Program
- Exit : Ferme SAS
- Save, Save as : sauvegarde le programme

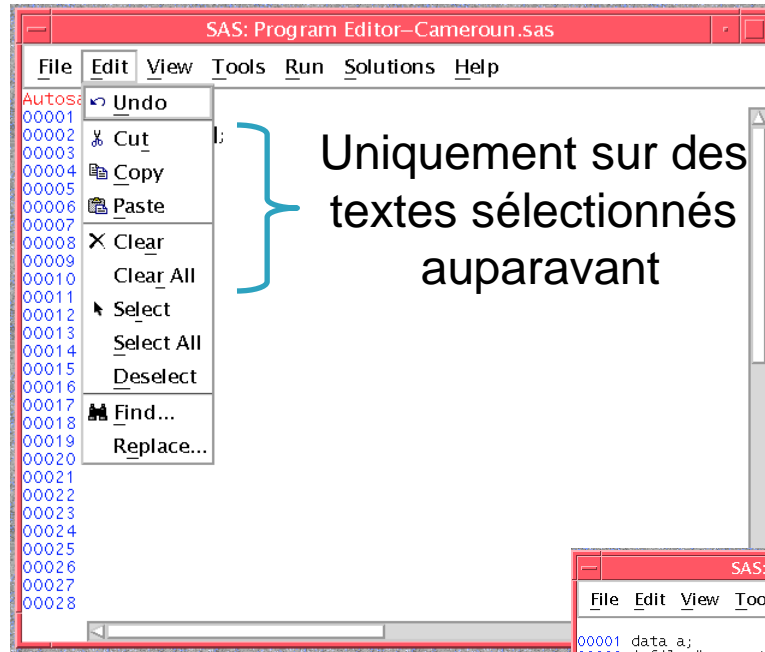
Dans la fenêtre « output », l'instruction « save as » permet de sauvegarder les résultats dans un fichier externe



Ne pas fermer sauvagement SAS, cela laisse le répertoire ouvert
Si un programme a planté, pensez à détruire le « core »



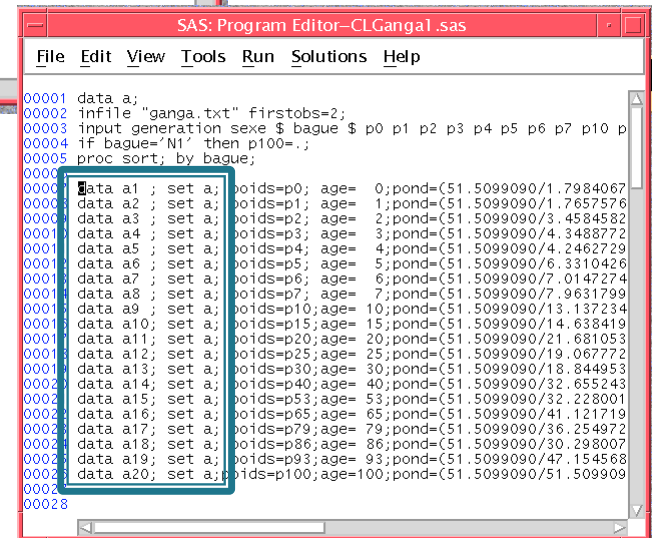
Le menu



Outil de sélection {
: a la sale manie de
se cacher derrière
les autres fenêtres {

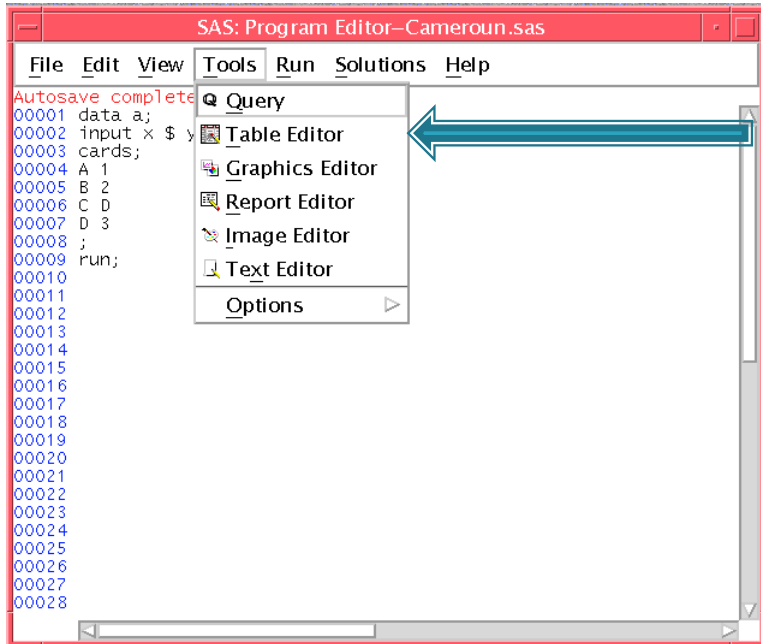


Touche Ctrl + Sélection :
Permet de sélectionner
seulement certaines colonnes

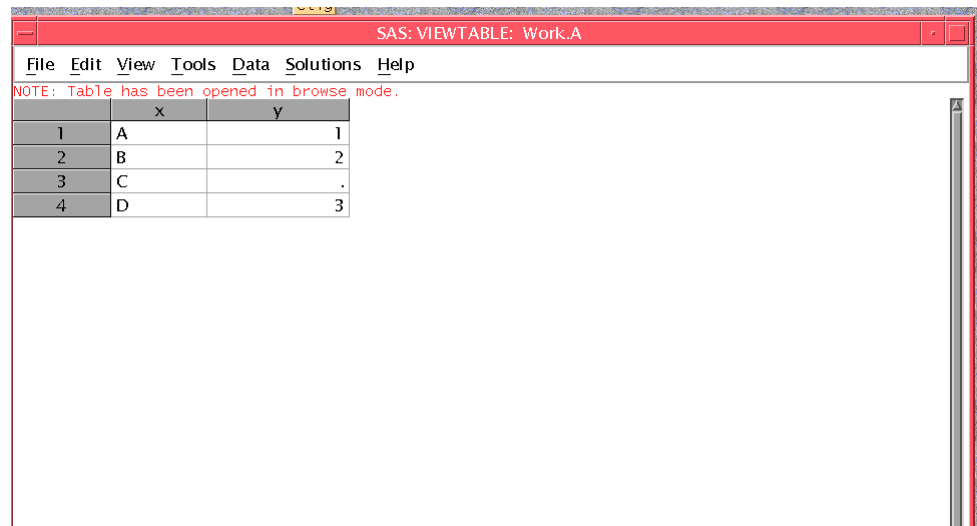




Le menu



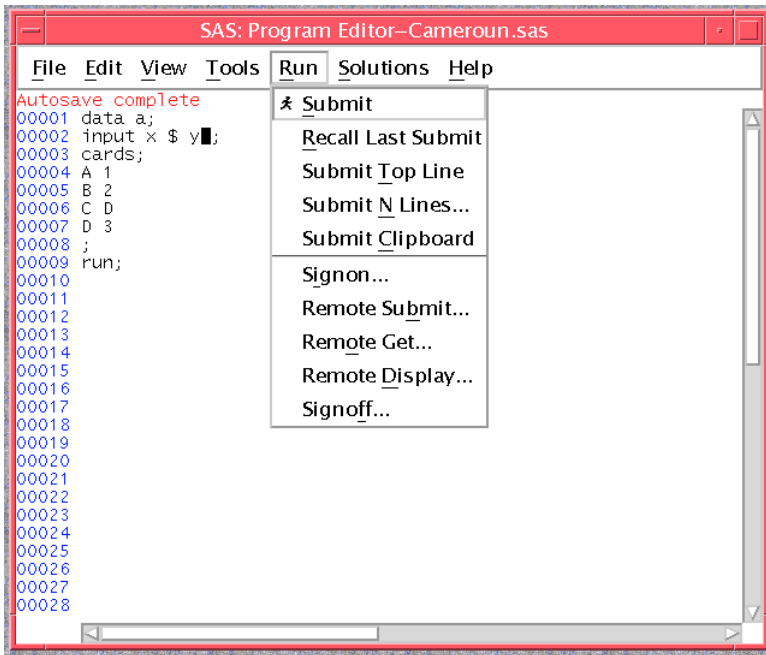
Permet de visualiser les fichiers temporaires créés pendant la session SAS



Fermer le fichier avant de retourner vers le programme, sinon SAS plante



Le menu



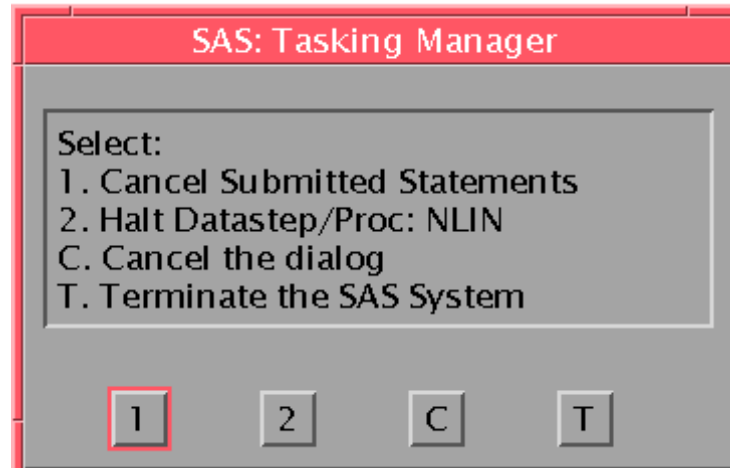
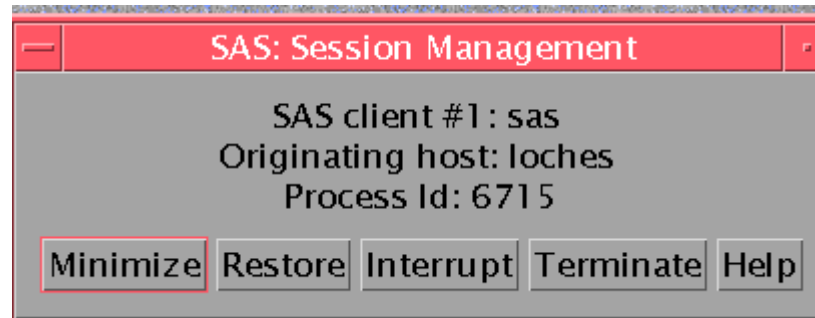
- ✚ **Submit** : lance le programme entier
- ✚ **Submit Top line** : lance le programme contenu dans la ligne 1
- ✚ **Submit N lines** : lance le programme contenu dans les N premières lignes
- ✚ **Submit Clipboard** : lance la partie de programme sélectionné

Dans le 1^{er} cas, le programme disparaît de l'écran.

Pour le rappeler : **Recall last submit**



En cas de problème avec le programme



Quelques raccourcis clavier



Ctrl + C : passe du mode insertion au mode délétion

Copier une ligne

Copier un bloc

Couper une ligne

Couper un bloc

Insérer une ligne

Insérer 6 lignes

Répéter une fois la ligne

Répéter une fois le bloc

Répéter 3 fois la ligne

Répéter 4 fois le bloc

Coller au dessous

Coller au dessus

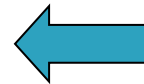
```
SAS: Program Editor--CLGangal .sas
File Edit View Tools Run Solutions Help

00001 data a;
00002 infile "ganga.txt" firstobs=2;
00003 input generation sexe $ bague $ p0 p1 p2 p3 p4 p5 p6 p7 p10 p15 p20
00004 if bague='N1' then p100=.;
00005 proc sort; by bague;
00006
C 1 data a1 ; set a; poids=p0; age= 0;pond=(51.5099090/1.7984067)**2;
00008 data a2 ; set a; poids=p1; age= 1;pond=(51.5099090/1.7657576)**2;
00009 data a3 ; set a; poids=p2; age= 2;pond=(51.5099090/3.4584582)**2;
CC data a4 ; set a; poids=p3; age= 3;pond=(51.5099090/4.3488772)**2;
00011 data a5 ; set a; poids=p4; age= 4;pond=(51.5099090/4.2462729)**2;
00012 data a6 ; set a; poids=p5; age= 5;pond=(51.5099090/6.3310426)**2;
CC data a7 ; set a; poids=p6; age= 6;pond=(51.5099090/7.0147274)**2;
00014 data a8 ; set a; poids=p7; age= 7;pond=(51.5099090/7.9631799)**2;
00015 data a9 ; set a; poids=p10;age= 10;pond=(51.5099090/13.1372346)**2;
00016 data a10; set a; poids=p15;age= 15;pond=(51.5099090/14.6384198)**2;
D 1 data a11; set a; poids=p20;age= 20;pond=(51.5099090/21.6810539)**2;
00018 data a12; set a; poids=p25;age= 25;pond=(51.5099090/19.0677721)**2;
00019 data a13; set a; poids=p30;age= 30;pond=(51.5099090/18.8449533)**2;
DD data a14; set a; poids=p40;age= 40;pond=(51.5099090/32.6552431)**2;
DD data a15; set a; poids=p53;age= 53;pond=(51.5099090/32.2280018)**2;
00022 data a16; set a; poids=p65;age= 65;pond=(51.5099090/41.1217197)**2;
00023 data a17; set a; poids=p79;age= 79;pond=(51.5099090/36.2549721)**2;
I 1 data a18; set a; poids=p86;age= 86;pond=(51.5099090/30.2980073)**2;
00025 data a19; set a; poids=p93;age= 93;pond=(51.5099090/47.1545685)**2;
00026 data a20; set a; poids=p100;age=100;pond=(51.5099090/51.5099090)**2;
00027
I 6
00029 data tout;
R 1 set a1 a2 a3 a4 a5 a8 a9 a10 a11 a13 a14 a15 a16 a17 a18 a19 a20;
00031
RR 1
00033
00034
00035
00036
00037
RR 1
00039
00040
R 3
00042
00043
00044
RR 4
00046 *****
00047 ***** MODELE DE GOMPERTZ *****
00048 *****
RR 4
00050
00051 proc nlin data=tout noprint;
A by bague;
00052 parms POC=10 to 30 by 5 B=0 to 100 by 10 K=0 to 10 by 1;
B *bounds 15<=POC<=23, B>0, K>0;
```

Mise en commentaire

- ▶ « * » met en commentaire tout ce qui suit, jusqu'au prochain point virgule

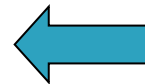
```
data a;  
infile « monfichier »;  
input mavARIABLE;  
* ceci est un commentaire ;  
proc means; var mavARIABLE;  
run;
```



Seule cette ligne est en commentaire, la procédure proc means est exécutée

- ▶ « /* ... */ » met en commentaire tout ce qui se trouve au milieu

```
data a;  
infile « monfichier »;  
input mavARIABLE;  
/* ceci est un commentaire ;  
proc means; var mavARIABLE; */  
run;
```

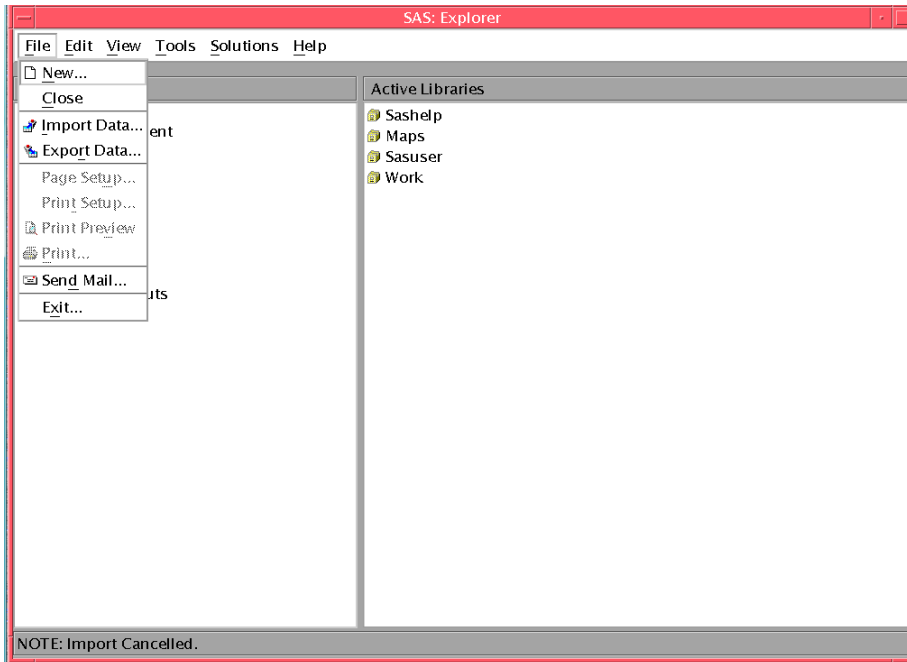


Les 2 lignes sont en commentaire, la procédure proc means n'est pas exécutée



L'étape DATA : lecture d'un fichier

On importe un fichier avec le menu « File/Import »



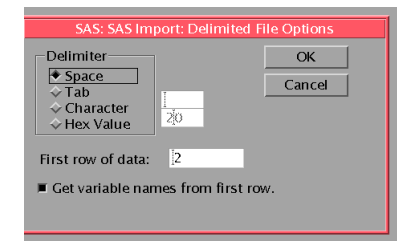
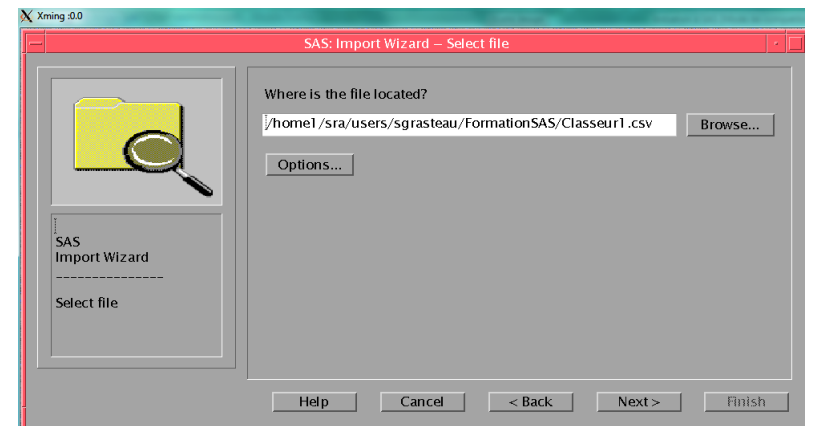
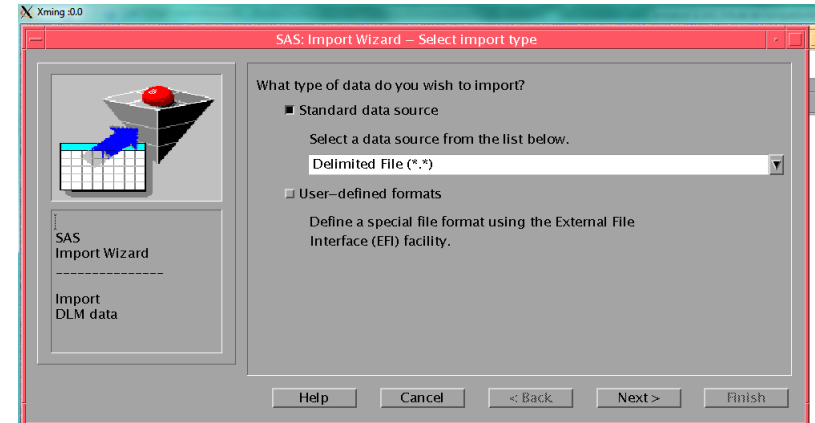
Ajouter une colonne vide à la fin du fichier à importer

Type de fichier : csv, txt, prn

Laisser le paramétrage « *Delimited file* »

Dans les options : Space pour prn, Tab pour txt; Character+ ; pour csv

Par défaut, le fichier se place dans le répertoire « Work »





L'étape DATA : lecture d'un fichier

On rentre les données directement dans le programme « cards »

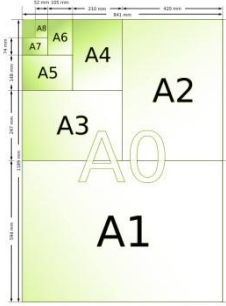
```
data a;
input individu date ddmmyy8. heure time8. categorie $ valeur 4.2;
format date ddmmyy8.;
format heure time8.;
aujourd'hui=today();
format aujourd'hui ddmmyy8.;
temps=aujourd'hui-date;
cards;
1 10022011 15:36 A 7.12
2 07112009 7:24 B 6.24
3 6031999 22:13 C 3.22
;
run;
proc print;
run;
```

On utilise l'instruction « infile » pour lire un fichier extérieur

```
data a;
infile "Exemple";
input individu date ddmmyy8. heure time8. categorie $ valeur 4.2;
format date ddmmyy8.;
format heure time8.;
aujourd'hui=today();
format aujourd'hui ddmmyy8.;
temps=aujourd'hui-date;
run;
proc print;
run;
```

Infile :

- Indique le chemin du fichier à lire
 - Ex : "FormationSAS/Fichier "
- **dlim** : spécifie quel caractère sépare les données
 - csv : *infile "Fichier " dlm=';';*
 - txt (séparateur tab) : *infile "Fichier " dlm='090D'x;*
- **Firstobs** : indique que la lecture commence à la ligne « n »
 - Pour commencer à lire à la ligne 4 : *infile "Fichier " firstobs=4;*



L'étape DATA : (in)formats

Input :

```
data a;
input individu date ddmmyy8. heure time8. categorie $ valeur 4.2;
format date ddmmyy8.;
format heure time8.;
aujourd'hui = today();
format aujourd'hui ddmmyy8.;
temps=aujourd'hui-date;
cards;
1 10022011 15:36 A 7.12
2 07112009 7:24 B 6.24
3 6031999 22:13 C 3.22
;
run;
proc print;
run;
```

- Nombres :

- Entier : Individu
- Réel : Valeur, écrit sur 4 positions dont 2 après la virgule

- Alphanumérique :

- Categorie : « \$ », lg. Max = 8 caractères
- Si plus de n caractères (n>8):
 - fichier aligné : *variable \$n.*
 - fichier non aligné : *variable : \$n.*

- Format pour les dates : ex. du 7 décembre 1999

- SAS le traduit en « 14585 » (nb. j. depuis le 01/01/1960)
- *datew.* : de 5 à 9 caractères
 - *date5.* ⇒ 07DEC
 - *date7.* ⇒ 07DEC99
 - *date9.* ⇒ 07DEC1999
- *ddmmyyw.* : de 2 à 10 caractères
 - *ddmmyy6.* ⇒ 071299
 - *ddmmyy8.* ⇒ 07/12/99
 - *ddmmyy10.* ⇒ 07/12/1999

- Format pour les heures : ex. de 7:35

- *Timew.d* : de 2 à 20 caractères dont d pour les dixièmes de secondes
 - *Time5.* ⇒ 7:35
 - *Time8.* ⇒ 7:35:00
 - *Time11.2* ⇒ 7:35:00.00

- @@ : indique que les données se succèdent sur une même ligne

```
input var1 var2 var3 @@;
```

A6	A4	A2
A7	A5	
A3	A0	A1
A1		

L'étape DATA : (in)formats



Erreur sur la ligne 3

The screenshot shows two SAS windows. The top window, 'SAS: Log-Untitled', displays the following log output:

```

205 data a;
206 infile "FormationSAS/pp";
207 input x $ y;
208 run;

NOTE: The infile "FormationSAS/pp" is:
      File Name=/home1/sra/users/sgrasteau/FormationSAS/pp,
      Owner Name=sgrasteau,Group Name=sra,
      Access Permission=rw-----,
      File Size (bytes)=16

NOTE: Invalid data for y in line 3 3-3.
RULE:  ---+---1---+---2---+---3---+---4---+---5---+---
3      C G 3
x=C y=. _ERROR_=1 _N_=3
NOTE: 4 records were read from the infile "FormationSAS/pp".
      The minimum record length was 3.
      The maximum record length was 3.
NOTE: The data set WORK.A has 4 observations and 2 variables.
NOTE: DATA statement used:
      real time      0.37 seconds
      cpu time       0.01 seconds
  
```

The bottom window, 'SAS: Program Editor-Untitled', shows the source code with a note:

```

NOTE: 4 Line(s) recalled.
00001 data a;
00002 infile "FormationSAS/pp";
00003 input x $ y;
00004 run;
00005
00006
00007
00008
00009
00010
00011
00012
  
```

On the right side of the log window, a 'Fenêtre' (Window) menu is open, showing options: Fenêtre, Editer, Options, and Aide. The 'Aide' option is highlighted, and an arrow points to it with the text 'Fichier à lire'.

Fichier à lire

Programme de lecture



L'étape DATA : sauvegarder des fichiers externes

- ▶ Les instructions « file » et « put » font le pendant en écriture de « infile » et « input » en lecture
- ▶ **CopieNonFormattee** : on remet les variables telles qu'elles étaient définies en lecture, avec toutes les décimales pour la nouvelle variable SEV_1000hab
- ▶ **CopieFormattee** : on peut changer le format des variables, ici le nom du département sera tronqué à 12 caractères, la variable SEV_1000hab sera écrite sur 6 positions, dont 2 décimales
- ▶ **CopieAlignee** : on précise à quelle colonne on veut commencer l'écriture de chaque variable avec l'instruction « @ », le fichier est aligné. Attention à penser à conserver des espaces entre variables.

```
00045 data seveso;
00046 infile "FormationSAS/DonneesSEVESO.csv" firstobs=2 dlm=';';
00047 input departement :$23. autorisation sevesoHR sevesoRM region :$13.
00048     pop2009 superficie taille :$10. densite densiteb :$19.;
00049 SEV_1000hab=1000*(autorisation+sevesoHR+sevesoRM)/pop2009;
00050
00051 file "FormationSAS/CopieNonFormattee";
00052 put departement autorisation taille densiteb SEV_1000hab;
00053
00054 file "FormationSAS/CopieFormattee";
00055 put departement :$12. autorisation taille :$10. densiteb :$19. SEV_1000hab 6.2;
00056
00057 file "FormationSAS/CopieAlignee";
00058 put @1 departement $23. @25 autorisation 4. @30 taille $10.
00059     @41 densiteb 19. @60 SEV_1000hab 6.2;
00060
00061 run;
```



L'étape DATA : sauvegarder des fichiers externes

CopieNonFormattee

```
Corrèze 226 Moyen Densité très faible 0.9386733417
Oise 541 Moyen Densité Moyenne 0.7176353914
Vosges 416 Moyen Densité Faible 1.1074721488
Eure-et-Loir 360 Moyen Densité Faible 0.8759110301
Seine-et-Marne 673 Moyen Densité Moyenne 0.5369398681
Manche 729 Moyen Densité Faible 1.4715802322
Charente 516 Moyen Densité Faible 1.5387634713
Var 635 Moyen Densité Moyenne 0.641097335
Deux-Sèvres 623 Moyen Densité Faible 1.7257484407
Aube 302 Grand Densité Faible 1.0486946075
Savoie 300 Grand Densité Faible 0.7642959742
Eure 464 Grand Densité Faible 0.8260452053
Hérault 455 Grand Densité Moyenne 0.4530308944
Orne 389 Grand Densité très faible 1.3476719225
Indre-et-Loire 412 Grand Densité Faible 0.7259918757
Aude 348 Grand Densité Faible 1.0279552281
Somme 547 Grand Densité Faible 0.9841118654
```

CopieFormattee

```
Paris 233 Petit Forte Densité 0.10
Hauts-de-Sei 949 Petit Forte Densité 0.61
Seine-Saint- 1053 Petit Forte Densité 0.70
Val-de-Marne 536 Petit Forte Densité 0.41
Territoire-d 87 Petit Densité Moyenne 0.65
Martinique 116 Petit Forte Densité 0.31
Val-d'Oise 355 Petit Forte Densité 0.31
Guadeloupe 113 Petit Densité Moyenne 0.29
Essonne 340 Petit Forte Densité 0.29
Yvelines 323 Petit Forte Densité 0.24
La Réunion 269 Petit Forte Densité 0.34
Rhône 814 Petit Forte Densité 0.50
Haut-Rhin 476 Petit Densité Moyenne 0.67
Vaucluse 287 Petit Densité Moyenne 0.54
Tarn-et-Garo 152 Petit Densité Faible 0.66
Corse-du-Sud 57 Petit Densité très faible 0.43
Pyrénées-Ori 188 Petit Densité Moyenne 0.43
Alpes-Mariti 301 Petit Forte Densité 0.28
```

CopieAlignee

```
Paris 233 Petit Forte Densité 0.10
Hauts-de-Seine 949 Petit Forte Densité 0.61
Seine-Saint-Denis 1053 Petit Forte Densité 0.70
Val-de-Marne 536 Petit Forte Densité 0.41
Territoire-de-Belfort 87 Petit Densité Moyenne 0.65
Martinique 116 Petit Forte Densité 0.31
Val-d'Oise 355 Petit Forte Densité 0.31
Guadeloupe 113 Petit Densité Moyenne 0.29
Essonne 340 Petit Forte Densité 0.29
Yvelines 323 Petit Forte Densité 0.24
La Réunion 269 Petit Forte Densité 0.34
Rhône 814 Petit Forte Densité 0.50
Haut-Rhin 476 Petit Densité Moyenne 0.67
Vaucluse 287 Petit Densité Moyenne 0.54
Tarn-et-Garonne 152 Petit Densité Faible 0.66
Corse-du-Sud 57 Petit Densité très faible 0.43
```



Créer / Modifier les variables

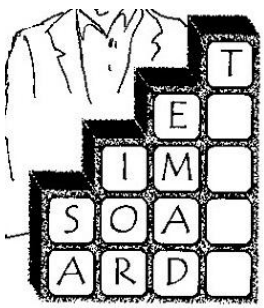
Opérations sur variables numériques : X=3.5 Y=10

Z=X*Y;	⇒	Z vaut 35
Z=min(X,Y);	⇒	Z vaut 3.5
Z=max(X,Y);	⇒	Z vaut 10
Z=Y**2;	⇒	Z vaut 100
Z=X**n;	⇒	Z vaut X puissance n (n positif ou négatif)
Z=sqrt(X);	⇒	Z vaut 1.87 (racine carrée de X, équivalent à X**0.5)
Z=log(X);	⇒	Z vaut 1.25 (log népérien)
Z=log10(X);	⇒	Z vaut 0.54 (log base 10)
Z=exp(X);	⇒	Z vaut 3.12
Z=mean(X,Y);	⇒	Z vaut 6.75
Z=sin(X) ;	⇒	Z vaut 0.06
Z=arcsin(X/10);	⇒	Z vaut 20.49

Pour modifier une variable déjà existante :

X+1; est équivalent à **X=X+1;**





Créer/Modifier les variables

Opérations sur variables alphanumériques : X='ceciestunessai', Y='MAJUSCULE'

Z=substr(X,5,3);	⇒	Z est la sous-chaîne de X, qui commence au caractère 5 et fait 3 caractères de long, soit 'est'
Z=lowcase(Y);	⇒	Z est la chaîne Y écrite en minuscule, soit 'majuscule'
Z=upcase(X);	⇒	Z est la chaîne X écrite en majuscule, soit 'CECIESTUNESSAI'
Z=X Y;	⇒	Z est la concaténation de X et Y, soit 'ceciestunessaiMAJUSCULE'
Z=compress(X,'e');	⇒	Z est la chaîne X sans les 'e', soit 'ccistunssai'

Quand on concatène des chaînes de caractères qui font moins de 8 caractères de long, SAS ajoute des blancs à la fin de la 1^{ère}, pour aller jusqu'à 8 :

```
X='abcd' ;  
Y='efgh' ;  
Z=X||Y;
```



Z='abcd efgh'

```
X='abcd' ;  
Y='efgh' ;  
Z=compress(X||Y, ' ');
```



Z='abcdefgh'



Créer/Modifier les variables

Opérations sur dates et heures

Today(); ⇒ Renvoie la date d'aujourd'hui

Time(); ⇒ Renvoie l'heure actuelle

X=Datdif(date_de_debut,date_de_fin,'Actual');

X=date_de_fin-date_de_debut;

Donne le nb de jours entre les dates de début et de fin

Y=heure_de_fin-heure_de_debut;

Donne le nb de minutes, secondes, dixièmes entre deux temps

```
SAS: Program Editor-Untitled
File Edit View Tools Run Solutions Help
NOTE: 16 Lines submitted.
00035
00036 data kkk;
00037 input debut ddmmyy8. fin ddmmyy9.;
00038 dif1=datdif(debut,fin,'Actual');
00039 dif2=fin-debut;
00040 ageactuel=(today()-debut);
00041 cards;
00042 01012000 10012000
00043 28012000 02022000
00044 31012000 02022000
00045 20022000 20032000
00046 20022001 20032001
00047 01032003 01032004
00048 01032004 01032005
00049 ;
00050 proc print;
00051 run;
00052
00053
```

```
SAS: Output-Untitled
File Edit View Tools Solutions Help
The SAS System 14:39 Tues
Obs debut fin dif1 dif2 ageactuel
1 14610 14619 9 9 4099
2 14637 14642 5 5 4072
3 14640 14642 2 2 4069
4 14660 14689 29 29 4049
5 15026 15054 28 28 3683
6 15765 16131 366 366 2944
7 16131 16496 365 365 2578
```

```
SAS: Program Editor-Untitled
File Edit View Tools Run Solutions Help
NOTE: 13 Lines submitted.
00054 data lll;
00055 input heureMAJ time5. heureABAT time6.;
00056 dureeMAJ=heureABAT-heureMAJ;
00057 delaiPRELEV=time()-heureABAT;
00058 format dureeMAJ time8.;
00059 format delaiPRELEV time11.2;
00060 cards;
00061 07:35 9:33
00062 07:37 9:36
00063 07:44 9:53
00064 ;
00065 proc print;
00066 run;
00067
00068
00069
```

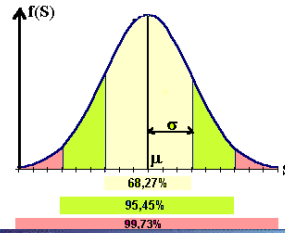
```
SAS: Output-Untitled
File Edit View Tools Solutions Help
The SAS System 14:
Obs heure MAJ heure ABAT dureeMAJ delaiPRELEV
1 27300 34380 1:58:00 1:01:17.17
2 27420 34560 1:59:00 0:58:17.17
3 27840 35580 2:09:00 0:41:17.17
```



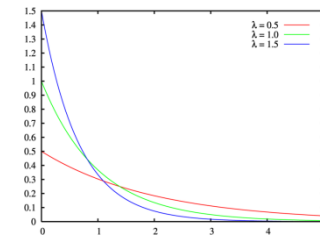

Tirages de variables aléatoires

Ces fonctions fournissent des nombres aléatoires issues de différentes distributions

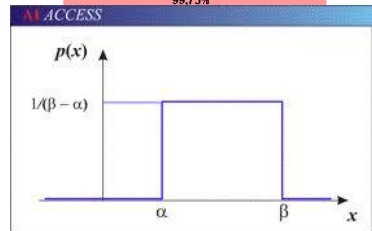
Loi normale : RANNOR(seed)



Loi exponentielle : RANEXP(seed)

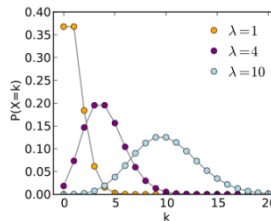


Loi uniforme : RANUNI(seed)



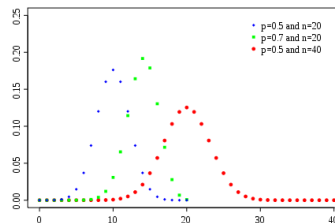
**Loi tabulée :
RANTBL(seed, p1, p2, ..., pn)**

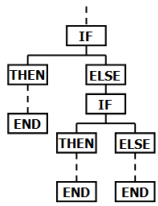
Loi de Poisson : RANPOI(seed)



L'utilisateur précise les probabilités associées à chacune des n catégories de la variable

Loi binomiale : RANBIN(seed)





If, then, else

1 seule instruction simple



```
if X>3 then Y=X+2;
```

Si X est plus grand que 3, Y vaudra X+2, sinon, SAS ne fait rien

```
if X='oui' then Y=1;
else Y=0;
```

Si X vaut 'oui', Y vaudra 1, sinon Y vaudra 0



Dans le 2^{ème} cas, Y vaut 0 même si X est une donnée manquante

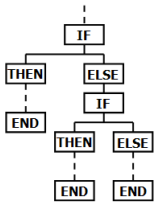
Plusieurs instructions



```
if X='oui' then do;
    Y=1;
    Z='peutetre';
end;
else if X='non' then do;
    Y=0;
    Z='non';
end;
else do;
    Y=.;
    Z='';
end;
```

Valeurs manquantes pour les variables numériques et alphanumériques





Opérateurs logiques pour les conditions

- ▶ =, NE, >, <, >=, <=
- ▶ Attention pour les variables numériques
 - Si X=. (donnée manquante) \Rightarrow X<n sera vrai
- ▶ AND, OR, NOT, IN
 - Pour exclure X=. d'une condition, il faut indiquer (X<n and X NE .)
 - Si X=4, if (X IN(2, 4, 6, 8)) sera vrai
- ▶ WHERE est équivalent à if
 - Par exemple if (X=4); est équivalent à where(X=4);



Retain

```

SAS: Program Editor-Untitled
File Edit View Tools Run Solutions Help
NOTE: 16 Lines submitted.
00068
00069 data prenom;
00070 input prenom:$12. sexe $;
00071 cards;
00072 MICHEL M
00073 YVES M
00074 ELISABETH F
00075 SOPHIE F
00076 DENIS M
00077 PHILIPPE M
00078 ;
00079
00080 data nombre1;set prenom;retain nb 0;nb+1;proc print;
00081 proc sort data=prenom; by sexe prenom;
00082 data nombre2; set prenom; by sexe;
00083 retain;if first.sexe then nb=0;nb+1;
00084 proc print;
00085 run;
00086

```

Permet de créer un compteur à incrémenter

- Sans condition (data nombre1)
- Avec condition (data nombre2)



Penser à initialiser le compteur



nombre1
(compteur sans condition)

SAS: Output-Untitled			
Tools Solutions Help			
The SAS System			
obs	prenom	sexe	nb
1	MICHEL	M	1
2	YVES	M	2
3	ELISABETH	F	3
4	SOPHIE	F	4
5	DENIS	M	5
6	PHILIPPE	M	6



nombre2
(compteur par sexe)

: Output-Untitled			
_help			
The SAS System			
obs	prenom	sexe	nb
1	ELISABETH	F	1
2	SOPHIE	F	2
3	DENIS	M	1
4	MICHEL	M	2
5	PHILIPPE	M	3
6	YVES	M	4



Drop, keep, delete

- ▶ Permet de garder ou de supprimer certaines variables ou observations
 - `Drop X;` ⇒ supprime la variable X
 - `Keep X Y;` ⇒ ne garde que les variables X et Y
 - `If X=1;` ⇒ ne garde que les observations pour lesquelles X=1
 - `If X=1 then delete;` ⇒ supprime les observations pour lesquelles X=1
 - `If _N_ < 100;` ⇒ ne garde que les observations 1 à 99



Exercice 1

- ▶ Lire le fichier « DonneesIndicePrixConsoEurope.csv »
 - Jour
 - Mois
 - Annee
 - Date
 - Indice des prix en France
 - Indice des prix en Grèce
 - Indice des prix au Royaume Uni
 - Indice des prix en Allemagne
 - Indice des prix en Pologne
 - Indice des prix dans l'UE
- ▶ Créer une nouvelle variable contenant le nom abrégé du mois sur 3 caractères
- ▶ Créer une variable contenant l'espace entre la date et le jour d'aujourd'hui
- ▶ Créer de nouvelles variables contenant la valeur de l'indice de chaque pays, rapporté à l'Union Européenne
- ▶ Créer une nouvelle variable valant « Pol>UK » si l'indice des prix relatif est supérieur en Pologne, « UK>Pol » dans le cas contraire
- ▶ Demander la fréquence des variables « mois », « mois sur 3 caractères », et « position relative de la Pologne et du Royaume uni », par « **proc freq; tables nom des variables** »
- ▶ Observer la variable « delai » pour vérifier qu'elle est bien calculée
- ▶ Observer la variable « position relative UK/Pologne » pour l'observation 1



Exercice 1

```
data prix;
infile "FormationSAS/Exercices/DonneesIndicesPrixConsoEurope.csv" dlm=';' firstobs=2;
input jour mois :$9. annee date ddmmyy10. france grece uk allemagne pologne UE;

mois3=substr(mois,1,3);
delai=today()-date;

France_rel=France/UE;
Grece_rel=Grece/UE;
Allemagne_rel=Allemagne/UE;
uk_rel=uk/UE;
pologne_rel=pologne/UE;

if pologne_rel>uk_rel then posrel='Pologne>UK';
    else posrel='UK>Pologne';

proc freq; tables mois mois3 posrel;

proc gplot; plot france_rel*date grece_rel*date uk_rel*date allemagne_rel*date pologne_rel*date/
overlay legend; symbol1 v=dot c=blue; symbol2 v=dot c=black; symbol3 v=dot c=green; symbol4
v=dot c=pink; symbol5 v=dot c=yellow;
run;
```



Exercice 1

The FREQ Procedure

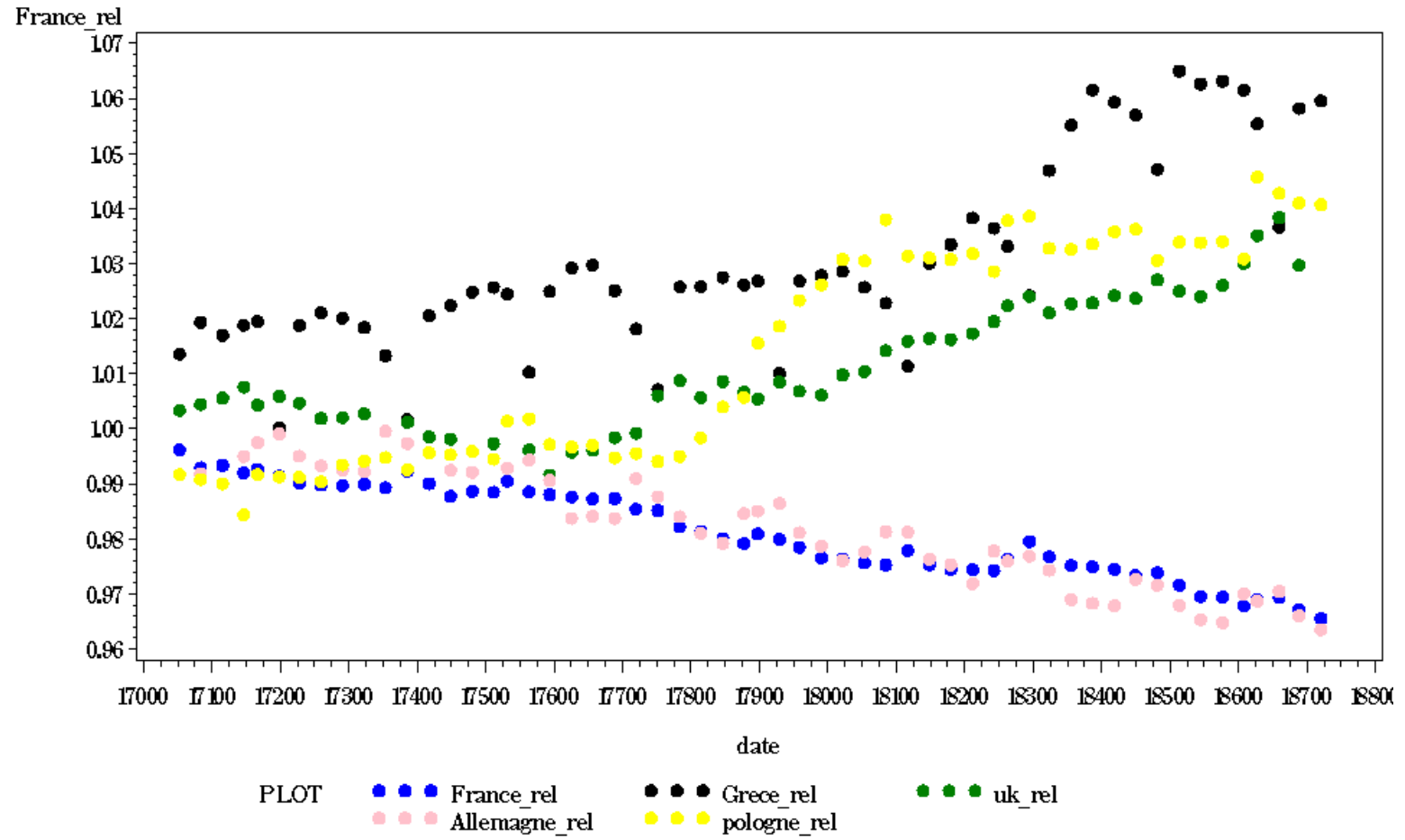
mois	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Août	4	7.14	4	7.14
Avril	5	8.93	9	16.07
Décembre	5	8.93	14	25.00
Février	5	8.93	19	33.93
Janvier	5	8.93	24	42.86
Juillet	4	7.14	28	50.00
Juin	4	7.14	32	57.14
Mai	4	7.14	36	64.29
Mars	5	8.93	41	73.21
Novembre	5	8.93	46	82.14
Octobre	5	8.93	51	91.07
Septembre	5	8.93	56	100.00

mois3	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Aoû	4	7.14	4	7.14
Avr	5	8.93	9	16.07
Déc	5	8.93	14	25.00
Fév	5	8.93	19	33.93
Jan	5	8.93	24	42.86
Jui	8	14.29	32	57.14
Mai	4	7.14	36	64.29
Mar	5	8.93	41	73.21
Nov	5	8.93	46	82.14
Oct	5	8.93	51	91.07
Sep	5	8.93	56	100.00

posrel	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Pologne>UK	33	58.93	33	58.93
UK>Pologne	23	41.07	56	100.00



Exercice 1





Exercice 1

SAS: VIEWTABLE: Work.Prix

File Edit View Tools Data Solutions Help

	jour	mois	annee	date	france	grece	uk	allemagne	pologne	UE	mois3	delai
1	4	Avril	2011	18721	111.33	122.17	.	111.1	120	115.3	Avr	74
2	3	Mars	2011	18689	110.93	121.37	118.1	110.8	119.4	114.69	Mar	106
3	2	Février	2011	18660	109.97	117.6	117.8	110.1	118.3	113.44	Fév	135
4	1	Janvier	2011	18628	109.44	119.19	116.9	109.4	118.1	112.93	Jan	167
5	12	Décembre	2010	18608	109.75	120.37	116.8	110	116.9	113.39	Déc	187
6	11	Novembre	2010	18577	109.22	119.78	115.6	108.7	116.5	112.66	Nov	218
7	10	Octobre	2010	18545	109.07	119.54	115.2	108.6	116.3	112.49	Oct	250
8	9	Septembre	2010	18514	108.91	119.38	114.9	108.5	115.9	112.09	Sep	281
9	8	Août	2010	18482	108.95	117.15	114.9	108.7	115.3	111.87	Aoû	313
10	7	Juillet	2010	18450	108.68	118.02	114.3	108.6	115.7	111.65	Jui	345
11	6	Juin	2010	18419	109.04	118.53	114.6	108.3	115.9	111.89	Jui	376
12	5	Mai	2010	18387	109.04	118.73	114.4	108.3	115.6	111.84	Mai	408
13	4	Avril	2010	18356	108.89	117.82	114.2	108.2	115.3	111.66	Avr	439
14	3	Mars	2010	18324	108.57	116.37	113.5	108.3	114.8	111.15	Mar	471
15	2	Février	2010	18295	107.99	112.91	112.9	107.7	114.5	110.24	Fév	500
16	1	Janvier	2010	18263	107.34	113.59	112.4	107.3	114.1	109.94	Jan	532
17	12	Décembre	2009	18243	107.6	114.47	112.6	108	113.6	110.44	Déc	552
18	11	Novembre	2009	18212	107.28	114.31	112	107	113.6	110.09	Nov	583
19	10	Octobre	2009	18180	107.11	113.59	111.7	107.2	113.3	109.91	Oct	615
20	9	Septembre	2009	18149	106.99	112.99	111.5	107.1	113.1	109.69	Sep	646
21	8	Août	2009	18117	107.23	110.9	111.4	107.6	113.1	109.65	Aoû	678
22	7	Juillet	2009	18085	106.64	111.84	110.9	107.3	113.5	109.34	Jui	710
23	6	Juin	2009	18054	107.18	112.68	111	107.4	113.2	109.85	Jui	741
24	5	Mai	2009	18022	107.03	112.76	110.7	107	113	109.62	Mai	773
25	4	Avril	2009	17991	106.87	112.48	110.1	107.1	112.3	109.43	Avr	804
26	3	Mars	2009	17959	106.71	111.98	109.8	107	111.6	109.05	Mar	836
27	2	Février	2009	17930	106.49	109.77	109.6	107.2	110.7	108.67	Fév	865
28	1	Janvier	2009	17898	106.05	111.02	108.7	106.5	109.8	108.11	Jan	897
29	12	Décembre	2008	17878	106.51	111.62	109.5	107.1	109.4	108.77	Déc	917
30	11	Novembre	2008	17847	106.79	111.96	109.9	106.7	109.4	108.96	Nov	948
31	10	Octobre	2008	17815	107.34	112.2	110	107.3	109.2	109.37	Oct	980
32	9	Septembre	2008	17784	107.4	112.16	110.3	107.6	108.8	109.34	Sep	1011



Exercice 1

SAS: VIEWTABLE: Work.Prix

File Edit View Tools Data Solutions Help

	uk	allemagne	pologne	UE	mois3	delai	France_rel	Grece_rel	Allemagne_rel	uk_rel	pologne_rel	posrel
1	.	111.1	120	115.3	Avr	74	0.9655680833	1.0595836947	0.9635732871	.	1.0407632264	Pologne>UK
2	118.1	110.8	119.4	114.69	Mar	106	0.9672159735	1.058243962	0.9660824832	1.0297323219	1.0410672247	Pologne>UK
3	117.8	110.1	118.3	113.44	Fév	135	0.9694111425	1.0366713681	0.9705571227	1.0384344147	1.042842031	Pologne>UK
4	116.9	109.4	118.1	112.93	Jan	167	0.9690959001	1.0554325688	0.9687416984	1.0351545205	1.045780572	Pologne>UK
5	116.8	110	116.9	113.39	Déc	187	0.9678984037	1.0615574566	0.9701031837	1.0300731987	1.0309551107	Pologne>UK
6	115.6	108.7	116.5	112.66	Nov	218	0.9694656489	1.0631990059	0.9648499911	1.0260962187	1.0340848571	Pologne>UK
7	115.2	108.6	116.3	112.49	Oct	250	0.9695972975	1.0626722375	0.9654191484	1.0240910303	1.0338696773	Pologne>UK
8	114.9	108.5	115.9	112.09	Sep	281	0.9716299402	1.0650370238	0.9679721652	1.0250691409	1.0339905433	Pologne>UK
9	114.9	108.7	115.3	111.87	Aoû	313	0.9738982748	1.0471976401	0.971663538	1.0270850094	1.0306605882	Pologne>UK
10	114.3	108.6	115.7	111.65	Jui	345	0.9733990148	1.0570532915	0.9726824899	1.0237348858	1.0362740708	Pologne>UK
11	114.6	108.3	115.9	111.89	Jui	376	0.9745285548	1.0593439986	0.9679149164	1.0242202163	1.0358387702	Pologne>UK
12	114.4	108.3	115.6	111.84	Mai	408	0.9749642346	1.0616058655	0.9683476395	1.0228898426	1.0336194564	Pologne>UK
13	114.2	108.2	115.3	111.66	Avr	439	0.9751925488	1.0551674727	0.9690130754	1.0227476267	1.0325989611	Pologne>UK
14	113.5	108.3	114.8	111.15	Mar	471	0.9767881242	1.0469635628	0.9743589744	1.0211426001	1.0328385065	Pologne>UK
15	112.9	107.7	114.5	110.24	Fév	500	0.9795898855	1.0242198839	0.9769593614	1.0241291727	1.0386429608	Pologne>UK
16	112.4	107.3	114.1	109.94	Jan	532	0.9763507368	1.0331999272	0.9759869019	1.0223758414	1.0378388212	Pologne>UK
17	112.6	108	113.6	110.44	Déc	552	0.9742846795	1.036490402	0.9779065556	1.0195581311	1.0286128214	Pologne>UK
18	112	107	113.6	110.09	Nov	583	0.9744754292	1.0383322736	0.9719320556	1.0173494414	1.0318830048	Pologne>UK
19	111.7	107.2	113.3	109.91	Oct	615	0.974524611	1.0334819398	0.9753434628	1.0162860522	1.0308434173	Pologne>UK
20	111.5	107.1	113.1	109.69	Sep	646	0.9753851764	1.0300847844	0.9763880026	1.0165010484	1.0310876105	Pologne>UK
21	111.4	107.6	113.1	109.65	Aoû	678	0.9779297766	1.0113999088	0.9813041496	1.0159598723	1.0314637483	Pologne>UK
22	110.9	107.3	113.5	109.34	Jui	710	0.9753063838	1.0228644595	0.9813426011	1.0142674227	1.0380464606	Pologne>UK
23	111	107.4	113.2	109.85	Jui	741	0.9756941284	1.0257624033	0.9776968594	1.0104688211	1.0304961311	Pologne>UK
24	110.7	107	113	109.62	Mai	773	0.9763729246	1.028644408	0.976099252	1.0098522167	1.0308337895	Pologne>UK
25	110.1	107.1	112.3	109.43	Avr	804	0.9766060495	1.0278716988	0.9787078498	1.0061226355	1.0262268117	Pologne>UK
26	109.8	107	111.6	109.05	Mar	836	0.9785419532	1.026868409	0.9812012838	1.0068775791	1.0233837689	Pologne>UK
27	109.6	107.2	110.7	108.67	Fév	865	0.9799392657	1.0101223889	0.9864728076	1.0085580197	1.0186804086	Pologne>UK
28	108.7	106.5	109.8	108.11	Jan	897	0.9809453335	1.026917029	0.9851077606	1.0054574045	1.0156322264	Pologne>UK
29	109.5	107.1	109.4	108.77	Déc	917	0.979222212	1.0262020778	0.9846465018	1.0067114094	1.0057920382	UK>Pologne
30	109.9	106.7	109.4	108.96	Nov	948	0.9800844347	1.0275330396	0.9792584435	1.0086270191	1.0040381791	UK>Pologne
31	110	107.3	109.2	109.37	Oct	980	0.9814391515	1.0258754686	0.9810734205	1.0057602633	0.9984456432	UK>Pologne
32	110.3	107.6	108.8	109.34	Sep	1011	0.9822571794	1.0257911103	0.9840863362	1.0087799524	0.9950612768	UK>Pologne

Trier les fichiers

Tri sur la variable « Equipe »

The screenshot shows the SAS Program Editor with the following code:

```
NOTE: 12 Lines submitted.
00090 data un;
00091 input prenom :$9. equipe $;
00092 cards;
00093 ANNE CM
00094 AGNES DYNNU
00095 SANDRINE SAQSE
00096 JOEL OEUF
00097 CLAUDETTE COMMUN
00098 ;
00099 proc sort; by equipe;
00100 proc print;
00101 run;
```

The SAS Output window displays the following table:

Obs	prenom	equipe
1	ANNE	CM
2	CLAUDETTE	COMMUN
3	AGNES	DYNNU
4	JOEL	OEUF
5	SANDRINE	SAQSE

Tri sur les variables « Equipe »,
Puis « prénom » intra-équipe

The screenshot shows the SAS Program Editor with the following code:

```
NOTE: 17 Lines submitted.
00090 data un;
00091 input prenom :$10. equipe $;
00092 cards;
00093 ANNE CM
00094 SOPHIER OEUF
00095 AGNES DYNNU
00096 SANDRINE SAQSE
00097 SOPHIE CM
00098 JOEL OEUF
00099 MARIE-JOSE COMMUN
00100 ELISABETH SAQSE
00101 IRENE DYNNU
00102 CLAUDETTE COMMUN
00103 ;
00104 proc sort; by equipe prenom;
00105 proc print;
00106 run;
```

The SAS Output window displays the following table:

Obs	prenom	equipe
1	ANNE	CM
2	SOPHIE	CM
3	CLAUDETTE	COMMUN
4	MARIE-JOSE	COMMUN
5	AGNES	DYNNU
6	IRENE	DYNNU
7	JOEL	OEUF
8	SOPHIER	OEUF
9	ELISABETH	SAQSE
10	SANDRINE	SAQSE

Avec l'option « descending », on trie dans l'ordre opposé

proc sort; by descending equipe prenom;

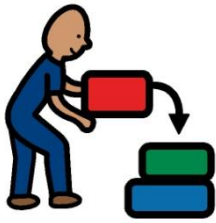


The screenshot shows the SAS Output window displaying the following table:

Obs	prenom	equipe
1	ELISABETH	SAQSE
2	SANDRINE	SAQSE
3	JOEL	OEUF
4	SOPHIER	OEUF
5	AGNES	DYNNU
6	IRENE	DYNNU
7	CLAUDETTE	COMMUN
8	MARIE-JOSE	COMMUN
9	ANNE	CM
10	SOPHIE	CM



Le fichier à trier est par défaut celui qui est en cours. On peut insérer l'instruction de tri à distance du fichier à trier, mais en ajoutant l'option « *proc sort data=fichier_a_trier ;* »



Gestion des fichiers : instruction set

SAS: Program Editor—Un

File Edit View Tools Run Solutions Help

```
NOTE: 22 Lines submitted.
00090 data un;
00091 input prenom :$9. equipe $;
00092 cards;
00093 ANNE CM
00094 AGNES DYNNU
00095 SANDRINE SAQSE
00096 JOEL OEUF
00097 CLAUDETTE COMMUN
00098 ;
00099
00100 data deux;
00101 input equipe $ etage;
00102 cards;
00103 CM 1
00104 COMMUN 0
00105 DYNNU 0
00106 SAQSE 1
00107 OEUF 2
00108 ;
00109
00110 data undeux;
00111 set un deux;
00112 proc print;
00113
00114 run;
00115
```

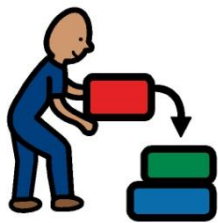
SAS: Output—Untitled

File Edit View Tools Solutions Help

The SAS System

Obs	prenom	equipe	etage
1	ANNE	CM	.
2	AGNES	DYNNU	.
3	SANDRINE	SAQSE	.
4	JOEL	OEUF	.
5	CLAUDETTE	COMMUN	.
6		CM	1
7		COMMUN	0
8		DYNNU	0
9		SAQSE	1
10		OEUF	2

L'instruction « set » empile les fichiers sans les fusionner.
Les variables présentes seulement dans le fichier B sont des données manquantes pour les observations venant du fichier A dans le fichier total.



Gestion des fichiers : instruction set-by

L'instruction « set-by » permet de lire un fichier ou de concaténer plusieurs fichier(s) trié(s)

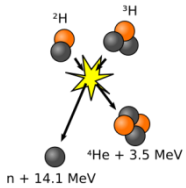
Ajouter l'instruction « **if first.cle_de_tri** » permet de ne garder que la 1^{ère} observation de chaque valeur de la clé de tri (if **last.cle_de_tri** garde la dernière). Utile en particulier pour récupérer automatiquement les résultats de procédures de calcul, qui génèrent souvent plusieurs lignes par individu.

```
SAS: Program Editor-Untitled
File Edit View Tools Run Solutions Help
00150 data tele;
00151 input nom $ numero;
00152 cards;
00153 TMC 10
00154 GULLI 18
00155 NRJ12 12
00156 ;
00157 proc sort; by numero;
00158
00159 data autres;
00160 input nom $ numero;
00161 cards;
00162 ZIDANE 10
00163 POMPIERS 18
00164 AVEYRON 12
00165 ;
00166 proc sort; by numero;
00167
00168
00169 data telautre;
00170 set tele autres; by numero;
00171 proc print;
00172 run;
00173
```

```
SAS: Output-Untitled
File Edit View Tools Solutions Help
The SAS System 14:39 T
Obs nom numero
1 TMC 10
2 ZIDANE 10
3 NRJ12 12
4 AVEYRON 12
5 GULLI 18
6 POMPIERS 18
```

```
SAS: Program Editor-Untitled
File Edit View Tools Run Solutions Help
00150 data tele;
00151 input nom $ numero;
00152 cards;
00153 TMC 10
00154 GULLI 18
00155 NRJ12 12
00156 ;
00157 proc sort; by numero;
00158
00159 data autres;
00160 input nom $ numero;
00161 cards;
00162 ZIDANE 10
00163 POMPIERS 18
00164 AVEYRON 12
00165 ;
00166 proc sort; by numero;
00167
00168
00169 data telautre;
00170 set tele autres; by numero;
00171 if first.numero;
00172 proc print;
00173 run;
```

```
SAS: Output-Untitled
File Edit View Tools Solutions Help
The SAS System 14:39 T
Obs nom numero
1 TMC 10
2 NRJ12 12
3 GULLI 18
```



Gestion des fichiers : instruction merge

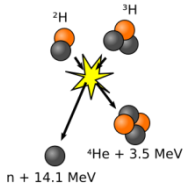
```
SAS: Program Editor–Untitled
File Edit View Tools Run Solutions Help
NOTE: 22 Lines submitted.
00090 data un;
00091 input prenom :$9. equipe $;
00092 cards;
00093 ANNE CM
00094 AGNES DYNNU
00095 SANDRINE SAQSE
00096 JOEL OEUF
00097 CLAUDETTE COMMUN
00098 ;
00099
00100 data deux;
00101 input equipe $ etage;
00102 cards;
00103 CM 1
00104 COMMUN 0
00105 DYNNU 0
00106 SAQSE 1
00107 OEUF 2
00108 ;
00109
00110 data undeux;
00111 merge un deux;
00112 proc print;
00113
00114 run;
00115
```

```
SAS: Output–Untitled
File Edit View Tools Solutions Help
The SAS System
Obs prenom equipe etage
1 ANNE CM 1
2 AGNES COMMUN 0
3 SANDRINE DYNNU 0
4 JOEL SAQSE 1
5 CLAUDETTE OEUF 2
```

L'instruction « merge » fusionne les fichiers ligne à ligne.



Le résultat n'est souvent pas celui attendu !



Gestion des fichiers : instruction merge-by

```

SAS: Program Editor-Untitled
File Edit View Tools Run Solutions Help
NOTE: 24 Lines submitted.
00090 data un;
00091 input prenom :$9. equipe $;
00092 cards;
00093 ANNE CM
00094 AGNES DYNNU
00095 SANDRINE SAQSE
00096 JOEL OEUF
00097 CLAUDETTE COMMUN
00098 ;
00099 proc sort; by equipe;
00100
00101 data deux;
00102 input equipe $ etage;
00103 cards;
00104 CM 1
00105 COMMUN 0
00106 DYNNU
00107 SAQSE 1
00108 OEUF 2
00109 ;
00110 proc sort; by equipe;
00111
00112 data undeux;
00113 merge un deux; by equipe;
00114 proc print;
00115 run;
  
```

SAS: Output-Untitled

The SAS System

obs	prenom	equipe	etage
1	ANNE	CM	1
2	CLAUDETTE	COMMUN	0
3	AGNES	DYNNU	0
4	JOEL	OEUF	2
5	SANDRINE	SAQSE	1

L'instruction « merge-by » fusionne les fichiers en fonction du clé de tri

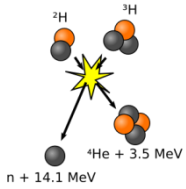


Le résultat est plus cohérent, mais cela implique de trier les fichiers auparavant !

Pour trier un fichier selon une variable utiliser **proc sort; by variable;**

Pour trier un fichier selon plusieurs variables utiliser **proc sort; by var1 var2;**

(le fichier sera alors trié d'abord sur var1, puis sur var2 intra-var1)



Gestion des fichiers : instruction merge-by

```
SAS: Program Editor-Untitled
File Edit View Tools Run Solutions Help
NOTE: 22 Lines submitted.
00117 data donnees;
00118 input poulet age poids;
00119 cards;
00120 1 35 1250
00121 2 33 1950
00122 3 34 2010
00123 4 35 752
00124 ;
00125 proc sort; by poulet;
00126
00127 data observations;
00128 input poulet obs $ ;
00129 cards;
00130 1 boiteux
00131 4 malade
00132 ;
00133 proc sort; by poulet;
00134
00135 data donneessansobs;
00136 merge donnees (in=i) observations (in=j);
00137 if i and (not j);
00138 by poulet;
00139 proc print;
00140 run;
00141
00142
00143
00144
00145
00146
00147
00148
00149
```

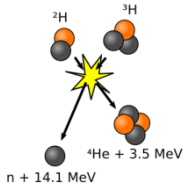
SAS: Output-Untitled

File Edit View Tools Solutions Help

The SAS System

obs	poulet	age	poids	obs
1	2	33	1950	
2	3	34	2010	

L'instruction « if i and (not j) » permet de garder les données présentes dans le fichier donnees, mais d'éliminer celles dans le fichier observations, donc d'éliminer les données aberrantes.



Gestion des fichiers : instruction merge-by

```
SAS: Program Editor-Untitled
File Edit View Tools Run Solutions Help
NOTE: 22 Lines submitted.
00117 data donnees;
00118 input poulet age poids;
00119 cards;
00120 1 35 1250
00121 2 33 1950
00122 3 34 2010
00123 4 35 752
00124 ;
00125 proc sort; by poulet;
00126 ;
00127 data observations;
00128 input poulet obs $ poids;
00129 cards;
00130 1 boiteux 1000
00131 4 malade 900
00132 ;
00133 proc sort; by poulet;
00134 ;
00135 data donneessansobs;
00136 merge donnees (in=i) observations (in=j);
00137 if i and (not j);
00138 by poulet;
00139 proc print;
00140 run;
00141 ;
00142 ;
00143 ;
00144 ;
00145 ;
00146 ;
00147 ;
00148 ;
00149 ;
```

SAS: Output-Untitled

The SAS System

obs	poulet	age	poids	obs
1	2	33	1950	
2	3	34	2010	

Si les deux fichiers à fusionner comprennent la même variable, les données du premier fichier écrasent celles du second



Exercice 2



🍌 Lire le fichier « DonneesCereales-Part1.csv ». Le fichier contient

- 🍌 numero : numéro de l'observation
- 🍌 calorie : quantité de calories/100 g de produit
- 🍌 protides : quantité de protides/100 g de produit
- 🍌 lipides : quantité de lipides/100 g de produit
- 🍌 glucides : quantité de glucides/100 g de produit
- 🍌 sucre : quantité de sucres rapides/100 g de produit
- 🍌 taux_cereale : proportion de céréales dans le produit
- 🍌 Sel : : quantité de sel/100 g de produit;

🍌 Lire le fichier « DonneesCereales-Part2.csv »

- 🍌 numero : numéro de l'observation
- 🍌 cat_cereale : catégorie de céréales (corn flakes, soufflee, fourree, muesli)
- 🍌 Parfum : parfum du produit (chocolat, fruits, nature, sucre)
- 🍌 cat_prix : categorie de prix (Premier prix, Distributeur, Marque)
- 🍌 enseigne : Enseigne distributrice (Auchan, Leclerc, Monoprix, Carrefour, Champion,'-')
- 🍌 cer_complete : présence de céréales complètes (oui, non)
- 🍌 Vit_Fer : enrichissement en vitamines et fer (oui, non)
- 🍌 nom : nom du produit (sur 45 caractères au maximum)

🍌 Comparer les résultats de :

- 🍌 Concaténation par « set »
- 🍌 Concaténation par « set by », par type de céréales
- 🍌 Fusion par merge



Exercice 2



🍌 Lire le fichier « DonneesCereales-Part1.csv »

```
data cereales1;  
infile "FormationSAS/Exercices/DonneesCereales-Part1.csv" dlm=';' firstobs=2;  
input numero calorie protides lipides glucides sucre taux_cereale sel;
```

🍌 Lire le fichier « DonneesCereales-Part2.csv »

```
data cereales2;  
infile "FormationSAS/Exercices/DonneesCereales-Part2.csv" dlm=';' firstobs=2;  
input numero cat_cereale :$11. parfum $ cat_prix :$12. enseigne :$12. cer_complete $ Vit_Fer $ nom :$45.;
```



Exercice 2



- **Comparer les résultats de :**

- **Concaténation par « set »**

```
data cerealesset;  
set cereales1 cereales2;
```

- **Concaténation par « set by », par type de céréales**

```
proc sort data=cereales1; by numero;  
proc sort data=cereales2; by numero;
```

```
data cerealessetby;  
set cereales1 cereales2; by numero;
```

- **Fusion par merge**

```
proc sort data=cereales1; by numero;  
proc sort data=cereales2; by numero;
```

```
data cerealesmerge;  
merge cereales1 cereales2; by numero;  
proc print;
```



Exercice 2



Instruction « SET » simple

Obs	numero	calorie	protides	lipides	glucides	sucree	taux-cereale	sel	cat-cereale	parfum	cat-prifix	enseigne	cecoocomplete	VitFer	nom
1	1	400	9.5	8.0	74.0	26.0	55	0.70							
2	2	370	10.0	1.5	79.0	23.0	63	0.70							
3	3	370	11.0	1.5	78.0	18.0	68	0.80							
4	4	354	10.0	6.0	65.0	23.0	.	0.45							
5	5	327	9.0	2.5	67.0	27.0	.	0.45							
6	6	366	8.0	0.7	82.0	9.0	87	0.80							
7	7	372	8.2	1.0	82.6	.	93	.							
8	8	363	8.1	0.8	80.8	.	93	.							
9	9	377	8.8	3.0	78.7	27.6	70	.							



Exercice 2



Instruction « SET BY »

obs	numéro	calories	protides	lipides	glucides	sucres	taux-cereale	sel	cat-cereale	parfum	cat-prix	enseigne	complete	Vit_Fer	nom
1	1	400	9.5	8.0	74	26	55	0.70	Corn Flakes	Chocolat	Distributeur	Champion	oui	non	Actiform choco petales
2	1
3	2	370	10.0	1.5	79	23	63	0.70	Corn Flakes	Fruits	Distributeur	Champion	oui	non	Actiform fruits petales
4	2
5	3	370	11.0	1.5	78	18	68	0.80	Corn Flakes	Nature	Distributeur	Champion	oui	non	Actiform nature petales
6	3
7	4	354	10.0	6.0	65	23	.	0.45	Corn Flakes	Chocolat	Marque	-	oui	oui	All Bran chocolat
8	4
9	5	327	9.0	2.5	67	27	.	0.45



Exercice 2

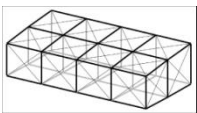


Instruction « MERGE BY »

.. ..

Obs	numero	calorie	protides	lipides	glucides	sucre	taux_cereale	sel	cat_cereale
1	1	400	9.5	8.0	74.0	26.0	55	0.70	Corn Flakes
2	2	370	10.0	1.5	79.0	23.0	63	0.70	Corn Flakes
3	3	370	11.0	1.5	78.0	18.0	68	0.80	Corn Flakes
4	4	354	10.0	6.0	65.0	23.0	.	0.45	Corn Flakes
5	5	327	9.0	2.5	67.0	27.0	.	0.45	Corn Flakes
6	6	366	8.0	0.7	82.0	9.0	87	0.80	Corn Flakes
7	7	372	8.2	1.0	82.6	.	93	.	Corn Flakes
8	8	363	8.1	0.8	80.8	.	93	.	Corn Flakes
9	9	377	8.8	3.0	78.7	27.6	70	.	Soufflee

Obs	parfum	cat_prix	enseigne	cer_complete	Vit_Fer	nom
1	Chocolat	Distributeur	Champion	oui	non	Actiform choco petales
2	Fruits	Distributeur	Champion	oui	non	Actiform fruits petales
3	Nature	Distributeur	Champion	oui	non	Actiform nature petales
4	Chocolat	Marque	-	oui	oui	All Bran chocolat
5	Fruits	Marque	-	oui	oui	All Bran Figue Pomme
6	Nature	Distributeur	Auchan	oui	oui	Auchan
7	Nature	Premier Prix	Auchan	oui	non	Auchan
8	Nature	Premier Prix	U	non	non	Bien Vu
9	Sucre	Distributeur	Carrefour	non	oui	Ble miel



L'étape PROC, structure générale

► **Proc nom_de_la_procedure data=fichier options;**

appelle la procédure, et stipule sur quel fichier travailler

dans les options, on trouve souvent ici le « noprint » qui supprime l'impression des résultats

by groupe;

si on veut exécuter la procédure par groupe de données (par sexe, génotype, ...)

ordre_specifique_a_la_procedure / options_de_l_ordre ;

l'ordre peut stipuler quelles variables sont concernées, quel est le modèle d'analyse, ...

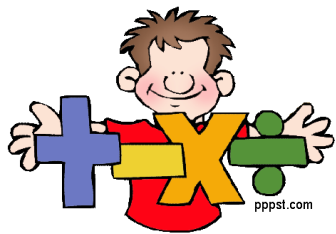
parfois, on stipule ici le nom de fichier de sortie

**output out=fichier_de_sortie
<caracteristiques_du_fichier_de_sortie> ;**

permet de sauvegarder les résultats de la procédure dans un fichier extérieur,

pas toujours à cet endroit là

Si on utilise cette option, le fichier « en cours » devient ce « fichier_de_sortie »



L'étape PROC, procédures de base

▶ Proc sort

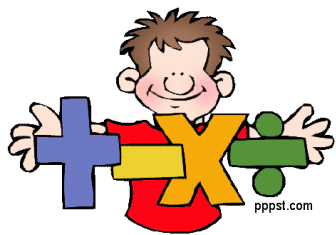
- Pour le tri
- Proc sort data=fichier ; by cle_de_tri1 cle_de_tri2 ...; par ordre croissant
- Proc sort data=fichier ; by descending cle_de_tri1 cle_de_tri2 ...; par ordre décroissant (pour la clé de tri 1), par ordre croissant pour la clé de tri 2

▶ Proc print

- Pour imprimer
- Proc print data=fichier; ⇒ imprime toutes les variables du fichier
- Proc print data=fichier; var v1 v2; ⇒ imprime seulement les variables v1, v2

▶ Proc contents

- Pour connaître le contenu du fichier (nom des variables et type des variables)
- Proc contents data=fichier;



L'étape PROC, procédures de base

The screenshot shows two windows from the SAS software. The top window, titled 'SAS: Program Editor--formationSASb.sas', contains the following SAS code:

```
NOTE: 12 Lines submitted.
00174 data tele2;
00175 input chaine $ pdm pdm_pub service_public $;
00176 cards;
00177 TF1      24.5  41.0 non
00178 France2 16.1  4.0  oui
00179 France3 10.7  2.0  oui
00180 Canal   3.1   2.0  non
00181 France5 4.7   0.5  oui
00182 M6     10.4  18.0 non
00183 TNT    19.7  23.0 non
00184 ;
00185 proc contents;run;
```

The bottom window, titled 'SAS: Output--Untitled', displays the output of the PROC CONTENTS procedure. It shows the following information:

The SAS System 15:12 Thursday, March 24, 2011

The CONTENTS Procedure

Data Set Name:	WORK.TELE2	Observations:	7
Member Type:	DATA	Variables:	4
Engine:	V8	Indexes:	0
Created:	15:13 Thursday, March 24, 2011	Observation Length:	32
Last Modified:	15:13 Thursday, March 24, 2011	Deleted observations:	0
Protection:		Compressed:	NO
Data Set Type:		Sorted:	NO
Label:			

-----Engine/Host Dependent Information-----

Data Set Page Size:	8192
Number of Data Set Pages:	1
First Data Page:	1
Max Obs per Page:	254
Obs in First Data Page:	7
Number of Data Set Repairs:	0
File Name:	/sra/users/sgrasteau/SAS_workE2000001933_1oches/tele2.sas7bdat
Release Created:	8.0101M0
Host Created:	SunOS
Inode Number:	2332046
Access Permission:	rw-----
Owner Name:	sgrasteau
File Size (bytes):	16384

-----Alphabetic List of Variables and Attributes-----

#	Variable	Type	Len	Pos
1	chaine	Char	8	16
2	pdm	Num	8	0
3	pdm_pub	Num	8	8
4	service_public	Char	8	24



ODS : mettre les résultats sous une jolie forme

Spécifie que le résultat sera sous format Word

```
ODS RTF FILE="Chemin/FichierSortie.doc" ;
```

Spécifie le nom du fichier de sortie

```
Proc machin;
```

```
run;
```

```
ODS RTF CLOSE;
```

C'est le résultat de la procédure « machin » qui sera dans FichierSortie.doc

Indispensable pour visualiser le fichier (créé au moment de l'instruction)

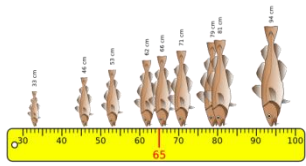
```
ODS HTML FILE="Chemin/GraphiqueSortie.doc" GPATH="Chemin/" (URL=NONE);
```

```
Proc gplot (gchart);
```

```
run; quit;
```

```
ODS HTML CLOSE;
```

Suppléments gratuits pour les graphiques



Statistiques élémentaires : proc means

```

SAS: Program Editor--Untitled
File Edit View Tools Run Solutions Help
NOTE: 20 Lines submitted.
00174 data tele2;
00175 input chaine $ pdm pdm_pub service_public $;
00176 cards;
00177 TF1      24.5  41.0 non
00178 France2 16.1  4.0  oui
00179 France3 10.7  2.0  oui
00180 Canal   3.1   2.0 non
00181 France5 4.7   0.5 oui
00182 M6     10.4  18.0 non
00183 TNT    19.7  23.0 non
00184 ;
00185
00186 proc means; var pdm ; title 'proc means de base';
00187
00188 proc means; class service_public; var pdm; title 'proc means par categorie';
00189
00190 proc means std stderr skewness kurtosis q1 median; var pdm_pub;
00191 title 'proc means avec des mots cles statistiques specifiques';
00192
00193 proc means noprint; var pdm pdm_pub;
00194 output out=ficmoyenne mean=mpdm mpdm_pub std=spdm spdm_pub;
00195
00196 proc print data=ficmoyenne;
00197 title 'proc means avec sauvegarde des resultats dans le fichier ficmoyenne';
00198
00199 run;
00200
00201
  
```

```

SAS: Output--Untitled
Proc Solutions Help
  
```

proc means de base 14:39 Tuesday, M
The MEANS Procedure
Analysis Variable : pdm

N	Mean	Std Dev	Minimum	Maximum
7	12.7428571	7.8021059	3.1000000	24.5000000

```

SAS: Output--Untitled
View Tools Solutions Help
  
```

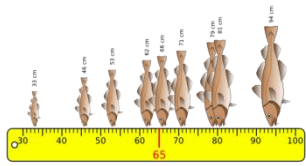
proc means par categorie 14:39 Tuesday, March 21
The MEANS Procedure
Analysis Variable : pdm

service_public	N Obs	N	Mean	Std Dev	Minimum	Maximum
non	4	4	14.4250000	9.5531408	3.1000000	24.5000000
oui	3	3	10.5000000	5.7026310	4.7000000	16.1000000



Faire un « **proc means; by categorie;** » donne le même résultat qu'un « **proc means; class categorie** », mais :

- La présentation diffère
- Pour le « by » il faut que le fichier soit trié par categorie avant, pas pour « class »



Statistiques élémentaires : proc means

```
SAS: Program Editor-Untitled
File Edit View Tools Run Solutions Help
NOTE: 20 Lines submitted.
00174 data tele2;
00175 input chaine $ pdm pdm_pub service_public $;
00176 cards;
00177 TF1      24.5  41.0 non
00178 France2  16.1   4.0 oui
00179 France3  10.7   2.0 oui
00180 Canal    3.1    2.0 non
00181 France5   4.7   0.5 oui
00182 M6      10.4  18.0 non
00183 TNT     19.7  23.0 non
00184 ;
00185
00186 proc means; var pdm ; title 'proc means de base';
00187
00188 proc means; class service_public; var pdm; title 'proc means par categorie';
00189
00190 proc means std stderr skewness kurtosis q1 median; var pdm_pub;
00191 title 'proc means avec des mots cles statistiques specifiques';
00192
00193 proc means noprint; var pdm pdm_pub;
00194 output out=ficmoyenne mean=mpdm mpdm_pub std=spdm spdm_pub;
00195
00196 proc print data=ficmoyenne;
00197 title 'proc means avec sauvegarde des resultats dans le fichier ficmoyenne';
00198
00199 run;
00200
00201
```

```
SAS: Output-Untitled
Edit View Tools Solutions Help
proc means avec des mots cles statistiques specifiques
14:39 Tuesday, March 22
```

The MEANS Procedure
Analysis Variable : pdm_pub

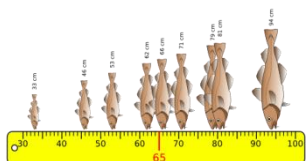
Std Dev	Std Error	Skewness	Kurtosis	Lower Quartile	Median
15.2108201	5.7491496	1.1954271	0.6207758	2.0000000	4.0000000

Ne produit aucun affichage dans la fenêtre « output »

```
SAS: Output-Untitled
Tools Solutions Help
proc means avec sauvegarde des resultats dans le fichier ficmoyenne
14:39 Tuesda:
Obs  _TYPE_  _FREQ_  mpdm  mpdm_pub  spdm  spdm_pub
1      0        7    12.7429  12.9286   7.80211  15.2108
```



Quand on a mis l'option « title », le titre s'affiche automatiquement sur toutes les pages qui vont suivre, jusqu'à ce que l'on mette un autre titre ou l'option **notitle**



Statistiques élémentaires : proc univariate

Pour l'essentiel, fait la même chose que proc means, mais donne par défaut plus d'information sur la distribution (quantiles, normalité de la distribution)

```
SAS: Program Editor-formationSASb.sas
File Edit View Tools Run Solutions Help
NOTE: 14 Lines submitted.
00203 data tele2;
00204 input chaine $ pdm pdm_pub service_public $;
00205 cards;
00206 TF1      24.5  41.0 non
00207 France2  16.1   4.0 oui
00208 France3  10.7   2.0 oui
00209 Canal    3.1    2.0 non
00210 France5   4.7   0.5 oui
00211 M6      10.4  18.0 non
00212 TNT     19.7  23.0 non
00213 ;
00214 proc univariate cibasic normal; var pdm;
00215 histogram pdm / normal(l=4 color=blue fill);
00216 run;
00217
```

Cibasic : donne l'intervalle de confiance de la moyenne

Normal : insère des statistiques sur la normalité de la distribution

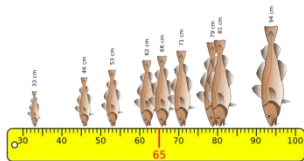
Histogram pdm : réalise un histogramme des valeurs de pdm

/normal : superpose une loi normale qui essaie de s'ajuster sur les données
(l=4 color=blue fill) cette courbe normale sera réalisée avec le trait n°4 (il y en a , l=1 trait plein ; l=3 trait pointillé) de couleur bleue, et l'aire sous la courbe est remplie

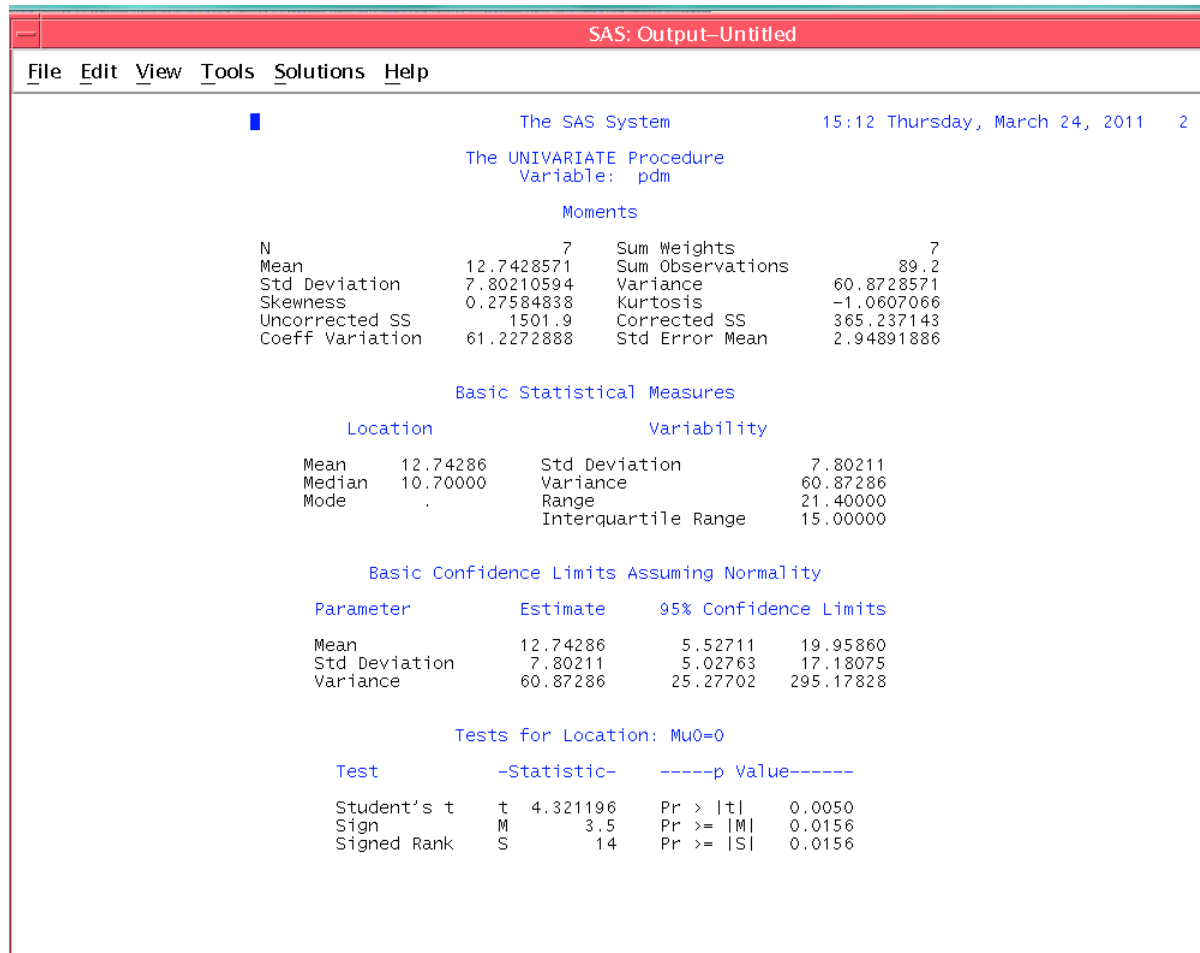


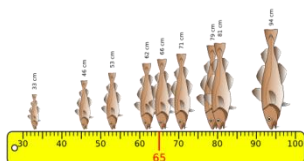
La plupart du temps « Kolmogorov-Smirnov » vous dit que votre distribution n'est pas normale.

Le test de Lilliefors est moins catégorique, mais il n'est pas inclus dans SAS (cf. commandement n°7, bureau 275)

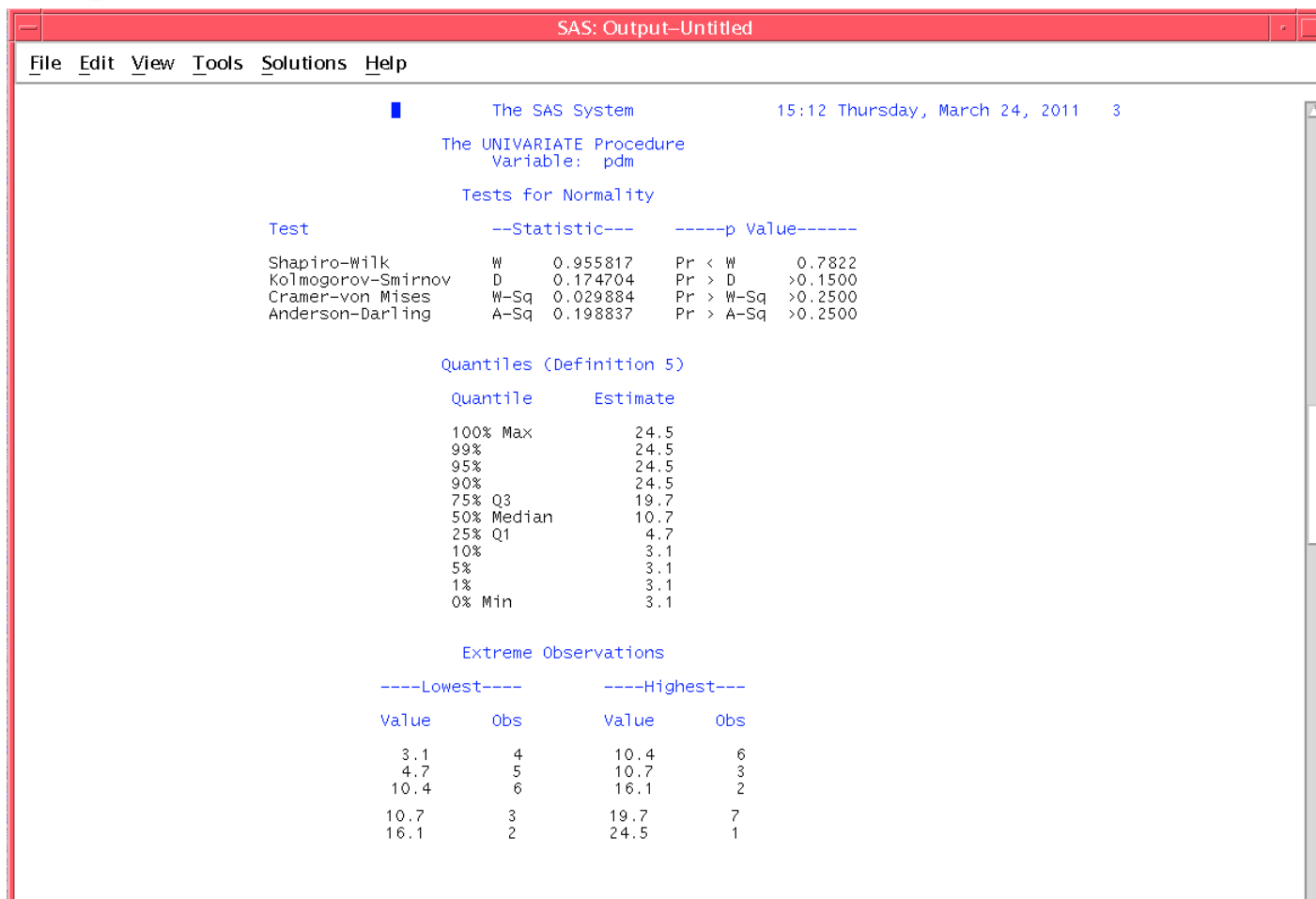


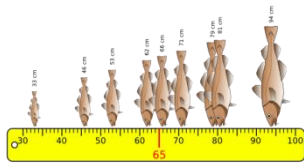
Statistiques élémentaires : proc univariate



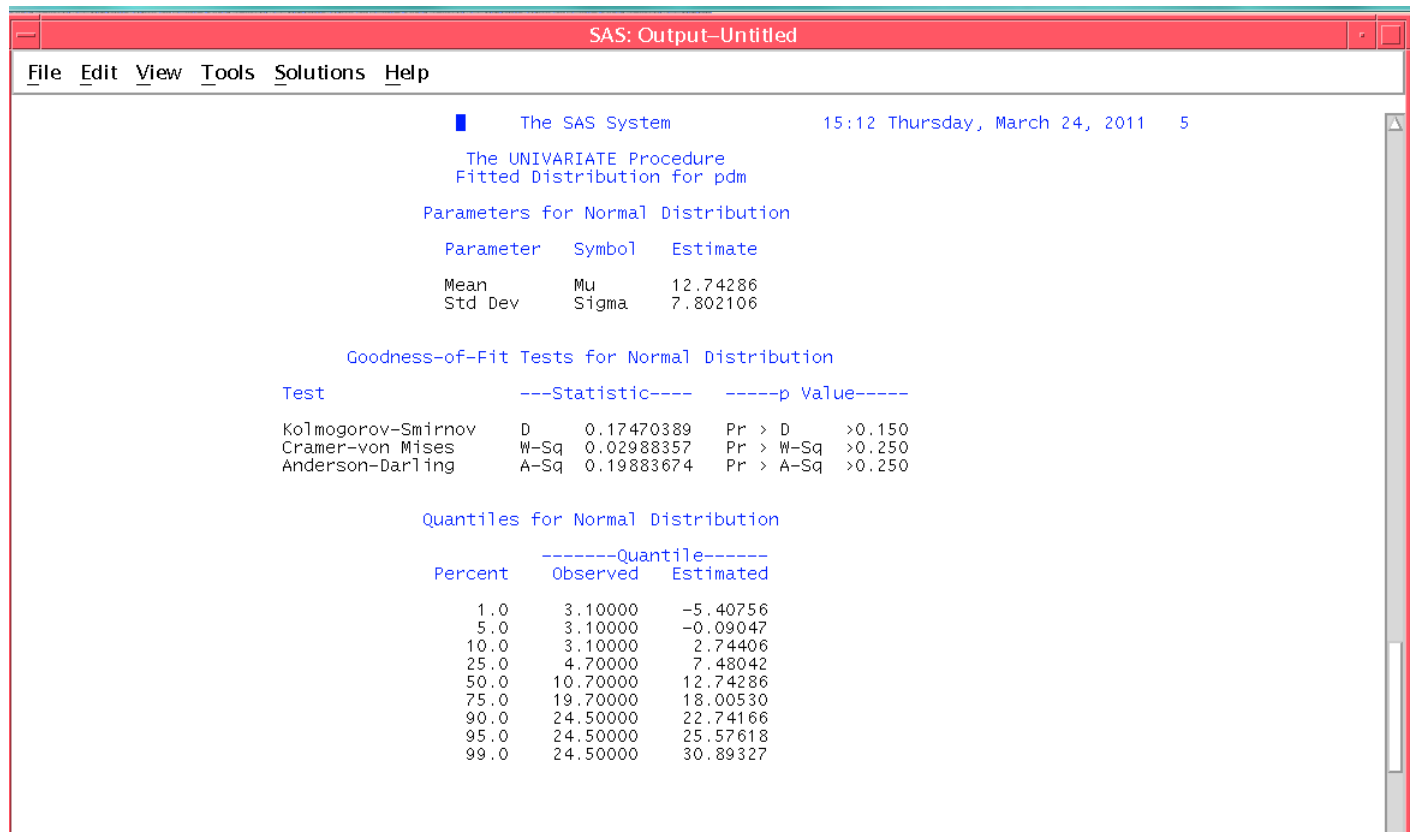


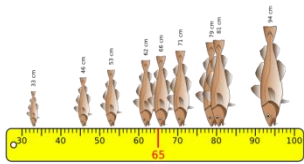
Statistiques élémentaires : proc univariate



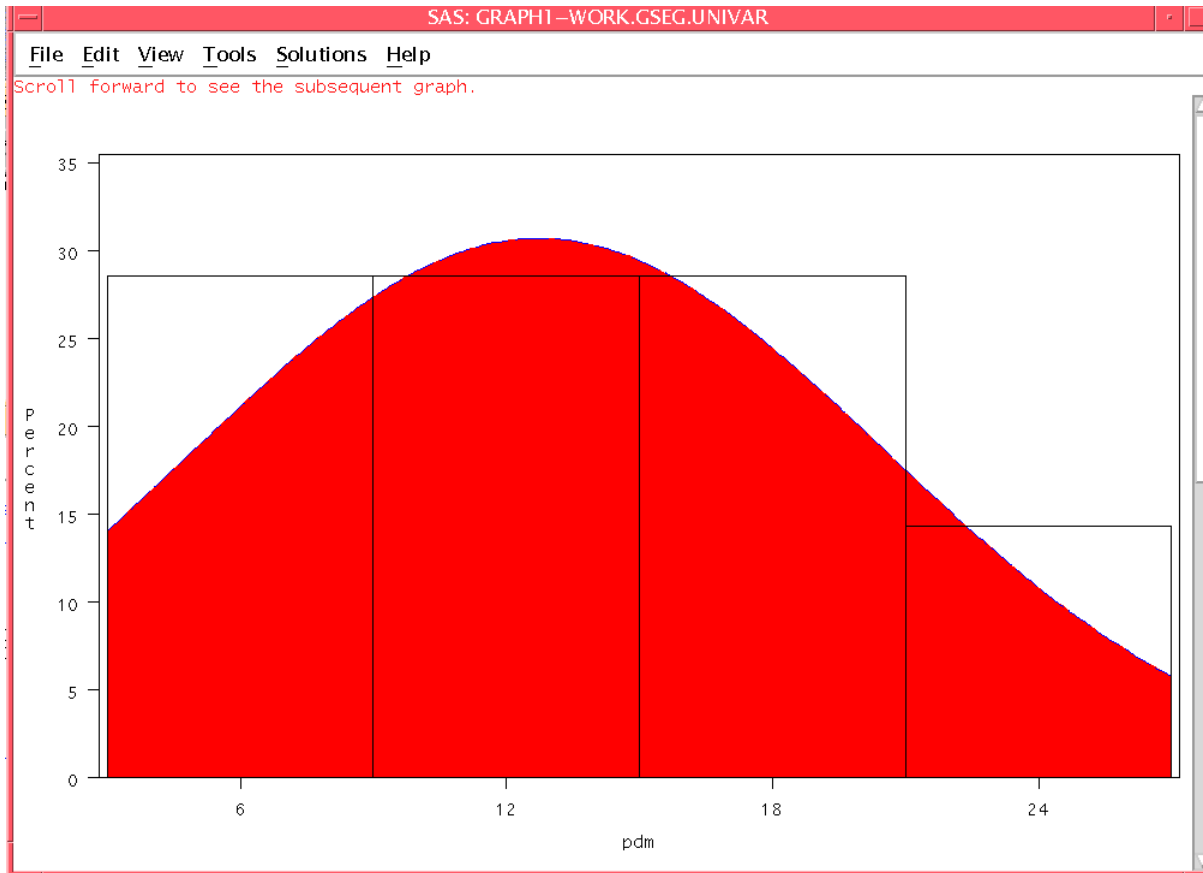


Statistiques élémentaires : proc univariate





Statistiques élémentaires : proc univariate





Exercice 2 suite



- Demander une moyenne générale des caractères numériques
- Demander une moyenne par catégorie de céréales :
 - avec l'option « by »
 - avec l'option « class »
- Demander une procédure univariante sur les variables lipides et taux de céréales :
 - Demander de tracer l'histogramme de la distribution
 - Demander la superposition d'une distribution normale



Exercice 2 suite



- Demander une moyenne générale des caractères numériques

```
proc means data=cerealesmerge;  
var calorie protides lipides glucides sucre taux_cereale sel;  
title 'Moyennes generales';
```

- Demander une moyenne par catégorie de céréales :
avec l'option « by »

```
proc sort data=cerealesmerge; by cat_cereale;  
proc means data=cerealesmerge; by cat_cereale;  
var calorie protides lipides glucides sucre taux_cereale sel;  
title 'Moyennes par categorie de cereales, option by';
```

avec l'option « class »

```
proc means data=cerealesmerge; class cat_cereale; var calorie protides lipides glucides sucre taux_cereale sel;  
title 'Moyennes par categorie de cereales, option class';
```



Exercice 2 suite



- Demander une moyenne générale des caractères numériques

Variable	N	Mean	Std Dev	Minimum	Maximum
calorie	113	389.1504425	31.9095321	321.0000000	483.0000000
protides	113	7.6026549	1.5507176	5.0000000	11.0000000
lipides	113	6.3017699	5.7595415	0.5000000	20.7000000
glucides	112	75.4571429	9.2646253	53.1000000	88.0000000
sucres	76	26.7052632	10.1940230	1.1000000	49.2000000
taux_cereale	78	63.8256410	17.0736482	31.0000000	100.0000000
sel	73	0.3897260	0.2553754	0.0100000	1.0000000



Exercice 2 suite



option « by »

option « class »

cat_cereale=Corn Flakes

cat_cereale	N	Variable	N	Mean	Std Dev	Minimum	Maximum
Corn Flakes	21	calorie	21	371.904761	15.709566	327.000000	410.000000
		protides	21	9	4	0	0
		lipides	21	8.4666667	1.4561364	5.5000000	11.0000000
		glucides	21	2.0142857	2.2913502	0.7000000	8.0000000
		sucre	16	79.9571429	5.8265403	65.0000000	87.7000000
		taux_cereal	17	18.5125000	8.2695324	8.0000000	30.1000000
		e	16	79.5882353	15.174233	55.0000000	98.0000000
		sel	16	0.6562500	2	0.3000000	0.8000000
Fourree	13	calorie	13	426.230769	16.669102	408.000000	470.000000
		protides	13	2	4	0	0
		lipides	13	7.2230769	0.7518354	6.0000000	8.5000000
		glucides	13	14.4076923	0.9142042	13.0000000	16.5000000
		sucre	8	66.6153846	2.7245324	63.0000000	72.0000000
		taux_cereal	12	35.1250000	3.1367636	32.0000000	42.0000000
		e	8	46.9166667	17.562658	31.0000000	98.0000000
		sel	8	0.2237500	4	0.1500000	0.3000000

Variable	N	Mean	Std Dev	Minimum	Maximum
calorie	21	371.9047619	15.7095664	327.0000000	410.0000000
protides	21	8.4666667	1.4561364	5.5000000	11.0000000
lipides	21	2.0142857	2.2913502	0.7000000	8.0000000
glucides	21	79.9571429	5.8265403	65.0000000	87.7000000
sucre	16	18.5125000	8.2695324	8.0000000	30.1000000
taux_cereale	17	79.5882353	15.1742332	55.0000000	98.0000000
sel	16	0.6562500	0.1815443	0.3000000	0.8000000

cat_cereale=Fourree

Variable	N	Mean	Std Dev	Minimum	Maximum
calorie	13	426.2307692	16.6691024	408.0000000	470.0000000
protides	13	7.2230769	0.7518354	6.0000000	8.5000000
lipides	13	14.4076923	0.9142042	13.0000000	16.5000000
glucides	13	66.6153846	2.7245324	63.0000000	72.0000000
sucre	8	35.1250000	3.1367636	32.0000000	42.0000000
taux_cereale	12	46.9166667	17.5626584	31.0000000	98.0000000
sel	8	0.2237500	0.0673875	0.1500000	0.3000000



Exercice 2 suite



- Demander une procédure univariate sur les variables lipides et le taux de céréales :

```
proc univariate data=cerealesmerge; var lipides taux_cereale ;  
histogram lipides taux_cereale / normal;  
title 'Distribution des variables taux de cereales et taux de lipides';
```




Exercice 2 suite



Taux de lipides

Moments			
N	113	Sum Weights	113
Mean	6.30176991	Sum Observations	712.1
Std Deviation	5.7595415	Variance	33.1723183
Skewness	0.92156323	Kurtosis	-0.5244978
Uncorrected SS	8202.79	Corrected SS	3715.29965
Coeff Variation	91.3956171	Std Error Mean	0.54181209

Tests for Location: $\mu_0=0$				
Test	Statistic		p Value	
Student's t	t	11.63091	Pr > t 	<.0001
Sign	M	56.5	Pr >= M 	<.0001
Signed Rank	S	3220.5	Pr >= S 	<.0001

Basic Statistical Measures			
Location		Variability	
Mean	6.301770	Std Deviation	5.75954
Median	3.500000	Variance	33.17232
Mode	1.000000	Range	20.20000
		Interquartile Range	8.00000

Quantiles (Definition 5)	
Quantile	Estimate
100% Max	20.7
99%	19.8
95%	18.0
90%	15.3
75% Q3	9.5
50% Median	3.5
25% Q1	1.5
10%	1.0
5%	0.8
1%	0.7
0% Min	0.5

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
0.5	109	18.0	112
0.7	19	18.4	52
0.7	6	19.0	61
0.8	86	19.8	55
0.8	20	20.7	51

NOTE: The mode displayed is the smallest of 2 modes with a count of 9.



Exercice 2 suite



Taux de céréales

Moments			
N	78	Sum Weights	78
Mean	63.825641	Sum Observations	4978.4
Std Deviation	17.0736482	Variance	291.509464
Skewness	0.4804631	Kurtosis	-0.3556679
Uncorrected SS	340195.8	Corrected SS	22446.2287
Coeff Variation	26.7504532	Std Error Mean	1.93320998

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	33.01537	Pr > t	<.0001
Sign	M	39	Pr >= M	<.0001
Signed Rank	S	1540.5	Pr >= S	<.0001

Basic Statistical Measures			
Location		Variability	
Mean	63.82564	Std Deviation	17.07365
Median	63.00000	Variance	291.50946
Mode	70.00000	Range	69.00000
		Interquartile Range	20.00000

Quantiles (Definition 5)	
Quantile	Estimate
100% Max	100
99%	100
95%	98
90%	93
75% Q3	70
50% Median	63
25% Q1	50
10%	41
5%	39
1%	31
0% Min	31

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
31	28	98	17
37	31	98	19
39	33	98	27
39	25	100	40
39	23	100	41

Missing Values			
Missing Value	Count	Percent Of	
		All Obs	Missing Obs
.	35	30.97	100.00



Exercice 2 suite



Taux de lipides

Parameters for Normal Distribution		
Parameter	Symbol	Estimate
Mean	Mu	6.30177
Std Dev	Sigma	5.759541

Goodness-of-Fit Tests for Normal Distribution				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.20878599	Pr > D	<0.010
Cramer-von Mises	W-Sq	1.27499156	Pr > W-Sq	<0.005
Anderson-Darling	A-Sq	7.26061567	Pr > A-Sq	<0.005

Quantiles for Normal Distribution		
Percent	Quantile	
	Observed	Estimated
1.0	0.70000	-7.09693
5.0	0.80000	-3.17183
10.0	1.00000	-1.07938
25.0	1.50000	2.41702
50.0	3.50000	6.30177
75.0	9.50000	10.18652
90.0	15.30000	13.68292
95.0	18.00000	15.77537
99.0	19.80000	19.70047

Taux de céréales

Parameters for Normal Distribution		
Parameter	Symbol	Estimate
Mean	Mu	63.82564
Std Dev	Sigma	17.07365

Goodness-of-Fit Tests for Normal Distribution				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.11522419	Pr > D	0.011
Cramer-von Mises	W-Sq	0.11834451	Pr > W-Sq	0.066
Anderson-Darling	A-Sq	0.91977571	Pr > A-Sq	0.020

Quantiles for Normal Distribution		
Percent	Quantile	
	Observed	Estimated
1.0	31.0000	24.1064
5.0	39.0000	35.7420
10.0	41.0000	41.9449
25.0	50.0000	52.3096
50.0	63.0000	63.8256
75.0	70.0000	75.3416
90.0	93.0000	85.7064
95.0	98.0000	91.9093
99.0	100.0000	103.5449

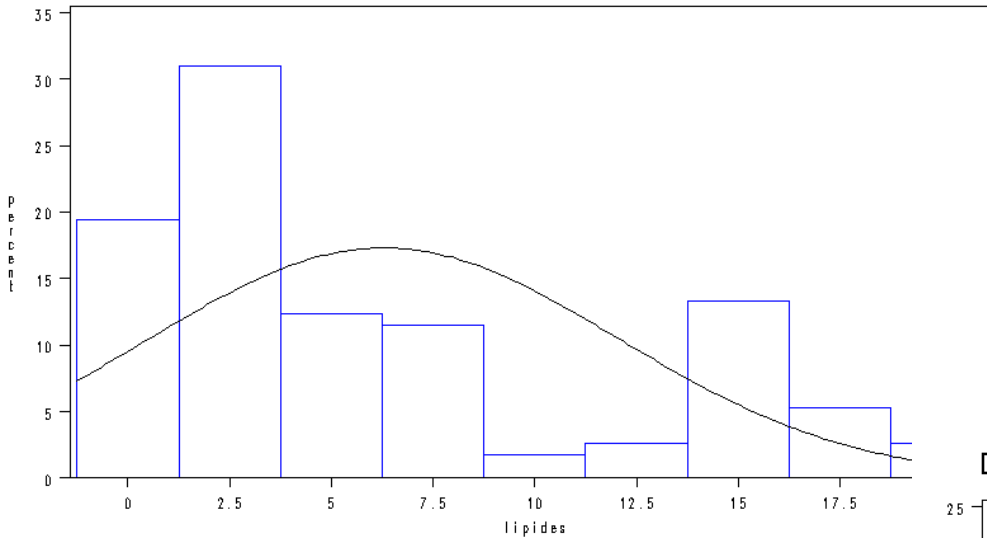


Exercice 2 suite



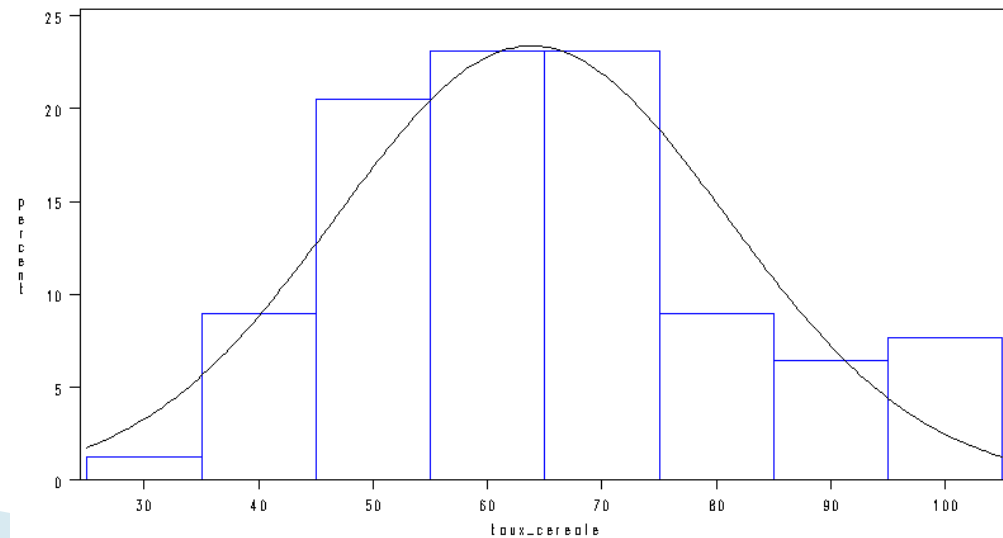
Lipides

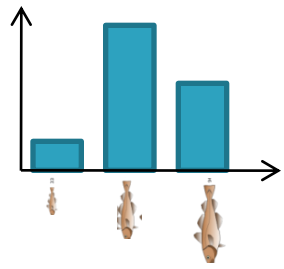
Distribution des variables taux de cereales et taux de lipides



Cereales

Distribution des variables taux de cereales et taux de lipides





Statistiques élémentaires : proc freq

SAS: Program Editor—Untitled

File Edit View Tools Run Solutions Help

```

Autosave complete
00216 run;
00217
00218
00219
00220
00221
00222 data un;
00223 input prenom :$10. equipe $ sexe;
00224 cards;
00225 ANNE CM 2
00226 SOPHIET CM 2
00227 MICHELD CM 1
00228 BERNARD CM 1
00229 ELISABETH CM 2
00230 NICOLE CM 2
00231 CHRISTOPHE CM 1
00232 PASCAL CM 1
00233 ESTELLEG CM 2
00234 ESTELLEA CM 2
00235 SABINE CM 2
00236 SONIA CM 2
00237 THIERRY CM 1
00238 JOEL OEUF 1
00239 YVES OEUF 1
00240 NICOLAS OEUF 1
00241 SOPHIER OEUF 2
00242 MAGALI OEUF 2
00243 JEAN-CLAUDE OEUF 1
00244 CEDRIC OEUF 1
00245 MARYSEM OEUF 2
00246 AURELIEN OEUF 1
00247 JEAN-MARC DYNNU 1
00248 NATHALIE DYNNU 2
00249 MICHEL DYNNU 1
00250 MARYSEL DYNNU 2
00251 AGNES DYNNU 2
00252 IRENE DYNNU 2
00253 SERGE DYNNU 1
00254 PHILIPPE DYNNU 1
00255 DANIEL DYNNU 1
00256 ANNE-MARIE DYNNU 2
00257 ELISABETH SAQSE 2
00258 FANNY SAQSE 2
00259 SANDRINE SAQSE 2
00260 MARIE SAQSE 2
00261 CHRISTELLE SAQSE 2
00262 SEVERINE SAQSE 2
00263 ;
00264
00265 proc freq; tables sexe equipe*sexe /chisq out=ficfreq;
00266 proc sort data=un; by equipe;
00267 proc print data=ficfreq;
00268 run;
00269
00270

```

SAS: Output—Untitled

Tools Solutions Help

The SAS System 15:12 Thur

The FREQ Procedure

sexe	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	16	42.11	16	42.11
2	22	57.89	38	100.00

Chi-Square Test for Equal Proportions

Chi-Square	0.9474
DF	1
Pr > ChiSq	0.3304

Sample Size = 38

The SAS System 15:1

Obs	equipe	sexe	COUNT	PERCENT
1	CM	1	5	13.1579
2	CM	2	8	21.0526
3	DYNNU	1	5	13.1579
4	DYNNU	2	5	13.1579
5	OEUF	1	6	15.7895
6	OEUF	2	3	7.8947
7	SAQSE	2	6	15.7895



Statistiques élémentaires : proc freq

Résultats du « tables sexe*equipe / chisq »

The SAS System 15:12 Thursday

The FREQ Procedure

Table of equipe by sexe

equipe	sexe		Total
Frequency	1	2	
Percent			
Row Pct			
Col Pct			
CM	5	8	13
	13.16	21.05	34.21
	38.46	61.54	
	31.25	36.36	
DYNNUT	5	5	10
	13.16	13.16	26.32
	50.00	50.00	
	31.25	22.73	
OEUF	6	3	9
	15.79	7.89	23.68
	66.67	33.33	
	37.50	13.64	
SAQSE	0	6	6
	0.00	15.79	15.79
	0.00	100.00	
	0.00	27.27	
Total	16	22	38
	42.11	57.89	100.00

SAS: Output-Untitled

Solutions Help

The SAS System 15:12 Thursday

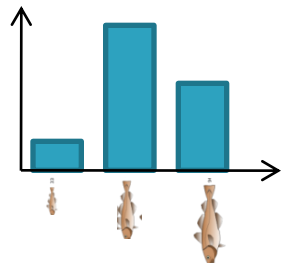
The FREQ Procedure

Statistics for Table of equipe by sexe

Statistic	DF	Value	Prob
Chi-Square	3	6.9174	0.0746
Likelihood Ratio Chi-Square	3	9.0844	0.0282
Mantel-Haenszel Chi-Square	1	0.5056	0.4771
Phi Coefficient		0.4267	
Contingency Coefficient		0.3924	
Cramer's V		0.4267	

WARNING: 50% of the cells have expected counts less than 5. Chi-Square may not be a valid test.

Sample Size = 38



Statistiques élémentaires : proc freq

- ▶ Options « **trend** » et « **relrisk** » pour :
 - Tester l'existence d'une tendance dans les données
 - Calculer les risques relatifs associés aux différentes modalités des variables

```
SAS: Program Editor-Untitled
File Edit View Tools Run Solutions Help
NOTE: 8 Line(s) recalled.
00001 data ronfle;
00002 infile "FormationSAS/DonneesRonflement.csv" firstobs=2 dlm=';';
00003 input patient $ age pere poids taille alcool sexe ronfle tabac;
00004
00005 proc freq; tables ronfle*alcool / trend;
00006
00007 proc freq; tables sexe*ronfle / relrisk nopercnt nocol;
```

Statistiques élémentaires : proc freq

The FREQ Procedure

Table of ronfle by alcool

ronfle		alcool						
		6	7	8	10	12	15	Total
Frequency	Percent							
Row Pct	Col Pct							
0		0	1	4	3	1	0	65
		0.00	1.00	4.00	3.00	1.00	0.00	65.00
		0.00	1.54	6.15	4.62	1.54	0.00	
		0.00	33.33	50.00	75.00	100.00	0.00	
1		2	2	4	1	0	1	35
		2.00	2.00	4.00	1.00	0.00	1.00	35.00
		5.71	5.71	11.43	2.86	0.00	2.86	
		100.00	66.67	50.00	25.00	0.00	100.00	
Total		2	3	8	4	1	1	100
		2.00	3.00	8.00	4.00	1.00	1.00	100.00

Statistics for Table of ronfle by alcool

Cochran-Armitage Trend Test

Statistic (Z)	-2.3641
One-sided Pr < Z	0.0090
Two-sided Pr > Z	0.0181

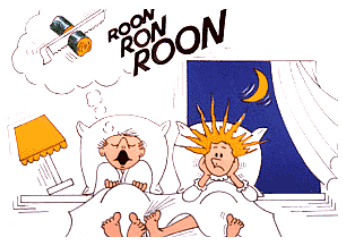
Sample Size = 100

H0 : pas de tendance linéaire entre les 2 variables

H1 « One Side » : tendance à la hausse (si $Z > 0$) ou à la baisse (si $Z < 0$)

H1 « Two-sided » : test bilatéral

Ici : plus on boit, moins on ronfle !



Statistiques élémentaires : proc freq

The SAS System

11:17 Monday

The FREQ Procedure

Table of sexe by ronfle

sexe	ronfle		
Frequency	0	1	Total
Row Pct			
0	45	30	75
	60.00	40.00	
1	20	5	25
	80.00	20.00	
Total	65	35	100

Statistics for Table of sexe by ronfle

Estimates of the Relative Risk (Row1/Row2)

Type of Study	Value	95% Confidence Limits	
Case-Control (Odds Ratio)	0.3750	0.1269	1.1080
Cohort (Col1 Risk)	0.7500	0.5729	0.9819
Cohort (Col2 Risk)	2.0000	0.8708	4.5937

Sample Size = 100

La cote du ronflement des individus de sexe « 0 » est de 0.375 par rapport à ceux de sexe « 1 »

Le risque de ne pas ronfler quand on est de sexe « 1 » est 0.75 fois plus élevé que si on est de sexe « 0 »

Le risque de ronfler quand on est de sexe « 1 » est 2 fois plus élevé que si on est de sexe « 0 »



Exercice 3

- ▶ Lire le fichier « `DonneesTachesMenageresContingence` » qui contient :
 - Tâche ménagère (jusqu'à 9 caractères)
 - Proportion des personnes y participant
 - Temps quotidien passé à cette tâche en min
 - Sexe

- ▶ Lire le fichier « `DonneesTachesMenageresBrut` » qui contient :
 - Sexe
 - Tâche ménagère (jusqu'à 9 caractères)
 - Participation (1 = 'oui'; 2='non')

- ▶ Réaliser un tableau de fréquence, en demandant un χ^2
 - Sur le fichier « brut »
 - Sur le fichier « contingence »



Exercice 3

```
data MENAGEContingence;  
infile "FormationSAS/Exercices/DonneesTachesMenageresContingence";  
input tache :$9. Prop Temps Sexe $;
```

```
proc freq; tables tache*sexe / chisq;  
weight prop;  
run;
```

```
data MENAGEBrut;  
infile "FormationSAS/Exercices/DonneesTachesMenageresBrut.csv" dlm=';' firstobs=2;  
input sexe $ tache :$9. participe;  
if participe=1;  
proc freq; tables sexe*tache / chisq;  
run;
```



Exercice 3

Repartition H/F des taches menageres

11:21 Friday, June 17, 2011 590

The FREQ Procedure

Table of sexe by tache

sexe	tache								
Frequency	Bricolage	Comptes	Courses	Cuisine	Enfants	Entretien	Linge	Menage	Total
Percent									
Row Pct									
Col Pct									
Femme	10	8	44	90	43	19	49	73	336
	1.85	1.48	8.12	16.61	7.93	3.51	9.04	13.47	61.99
	2.98	2.38	13.10	26.79	12.80	5.65	14.58	21.73	
	24.39	50.00	58.67	64.29	60.56	43.18	83.05	76.04	
Homme	31	8	31	50	28	25	10	23	206
	5.72	1.48	5.72	9.23	5.17	4.61	1.85	4.24	38.01
	15.05	3.88	15.05	24.27	13.59	12.14	4.85	11.17	
	75.61	50.00	41.33	35.71	39.44	56.82	16.95	23.96	
Total	41	16	75	140	71	44	59	96	542
	7.56	2.95	13.84	25.83	13.10	8.12	10.89	17.71	100.00

Statistics for Table of sexe by tache

Statistic	DF	Value	Prob
Chi-Square	7	52.0607	<.0001
Likelihood Ratio Chi-Square	7	53.1263	<.0001
Mantel-Haenszel Chi-Square	1	28.3767	<.0001
Phi Coefficient		0.3099	
Contingency Coefficient		0.2960	
Cramer's V		0.3099	

Sample Size = 542

$$X_{CR} = \frac{X - \mu_X}{\sigma_X} \sim \mathcal{N}(0,1)$$

Standardiser des données : proc standard

The screenshot shows two windows from the SAS interface:

- SAS: Program Editor-formationSASb.sas**: Contains the following SAS code:


```

Autosave complete
00274 data tele2;
00275 input chaine $ pdm pdm_pub service_public $
00276 cards;
00277 TF1      24.5  41.0 non
00278 France2  16.1   4.0 oui
00279 France3  10.7   2.0 oui
00280 Canal    3.1    2.0 non
00281 France5  4.7    0.5 oui
00282 M6       10.4  18.0 non
00283 TNT      19.7  23.0 non
00284 ;
00285 proc standard mean=0 std=1 out=std_tele;
00286 var pdm pdm_pub;
00287
00288 proc print data=std_tele;
00289 run;
00290
00291
00292
00293
00294
00295
      
```
- SAS: Output-Untitled**: Shows the output of the PROC PRINT statement, displaying the standardized data:

Obs	chaine	pdm	pdm_pub	service_public
1	TF1	1.50692	1.84549	non
2	France2	0.43029	-0.58699	oui
3	France3	-0.26183	-0.71847	oui
4	Canal	-1.23593	-0.71847	non
5	France5	-1.03086	-0.81709	oui
6	M6	-0.30029	0.33341	non
7	TNT	0.89170	0.66212	non

- **Mean** et **std** précisent quelle valeur est souhaitée pour standardiser (0 et 1 pour centrer réduire)
- Les valeurs sont sauvegardées sous le même nom de variable dans std_tele
- Si on veut avoir les valeurs brutes et les valeurs centrées-réduites, il faut créer une variable supplémentaire « pdmcr=pdm », puis appliquer proc standard sur pdmcr. On aura alors dans std_tele la variable brute (pdm) et la variable centrée réduite (pdmcr)



Graphiques sous SAS

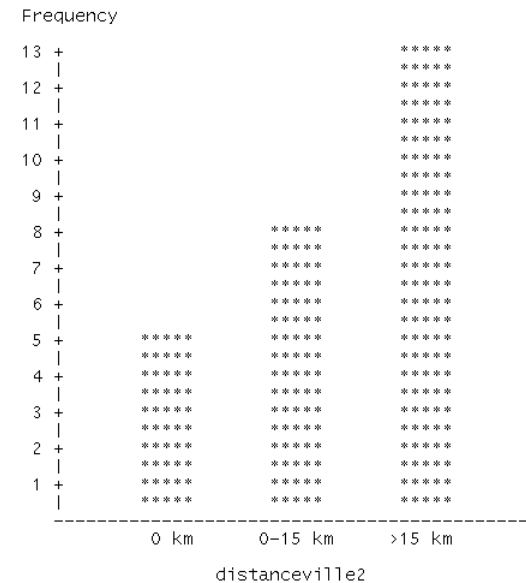


Il y a longtemps, les graphiques avec SAS, c'était ça :

```
data a;
  infile "FormationSAS/DonneesPopulation.csv" firstobs=2 dlm=';';
  input commune : $23. dpt cat $ distanceville distanceville2 $
  ;   celib marie veuf divorce total hommes femmes
  ;   enfants jeunes adultes TroisAge QuatAge
  ;   emploi chomeur militaire retraite etudiant autreinactif
  ;   maison_ind immeuble maison_retr hotel fortune piece autrelog;
proc chart; vbar distanceville2;run;
```

The SAS System

15:14 Thur





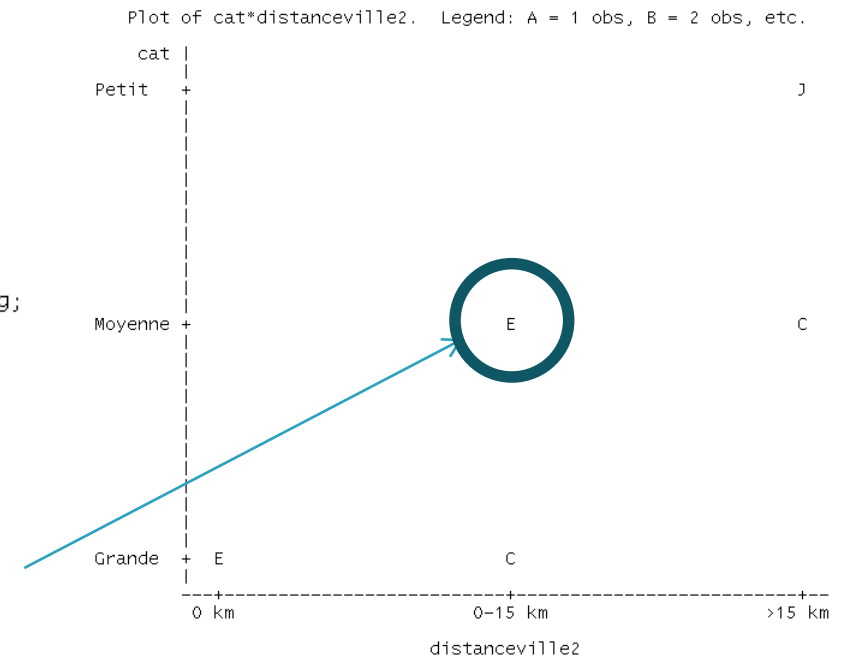
Graphiques sous SAS



Il y a longtemps, les graphiques avec SAS, c'était ça :

```
data a;
infile "FormationSAS/DonneesPopulation.csv" firstobs=2 dlm=';';
input commune : $23. dpt cat $ distanceville distanceville2 $
      celib marie veuf divorce total hommes femmes
      enfants jeunes adultes TroisAge QuatAge
      emploi chomeur militaire retraite etudiant autreinactif
      maison_ind immeuble maison_retr hotel fortune piece autrelog;
proc plot; plot cat*distanceville2;
run;
```

Plus la lettre est avancée dans l'alphabet, plus il y a de données au même endroit (A=1; B=2; C=3; ...)



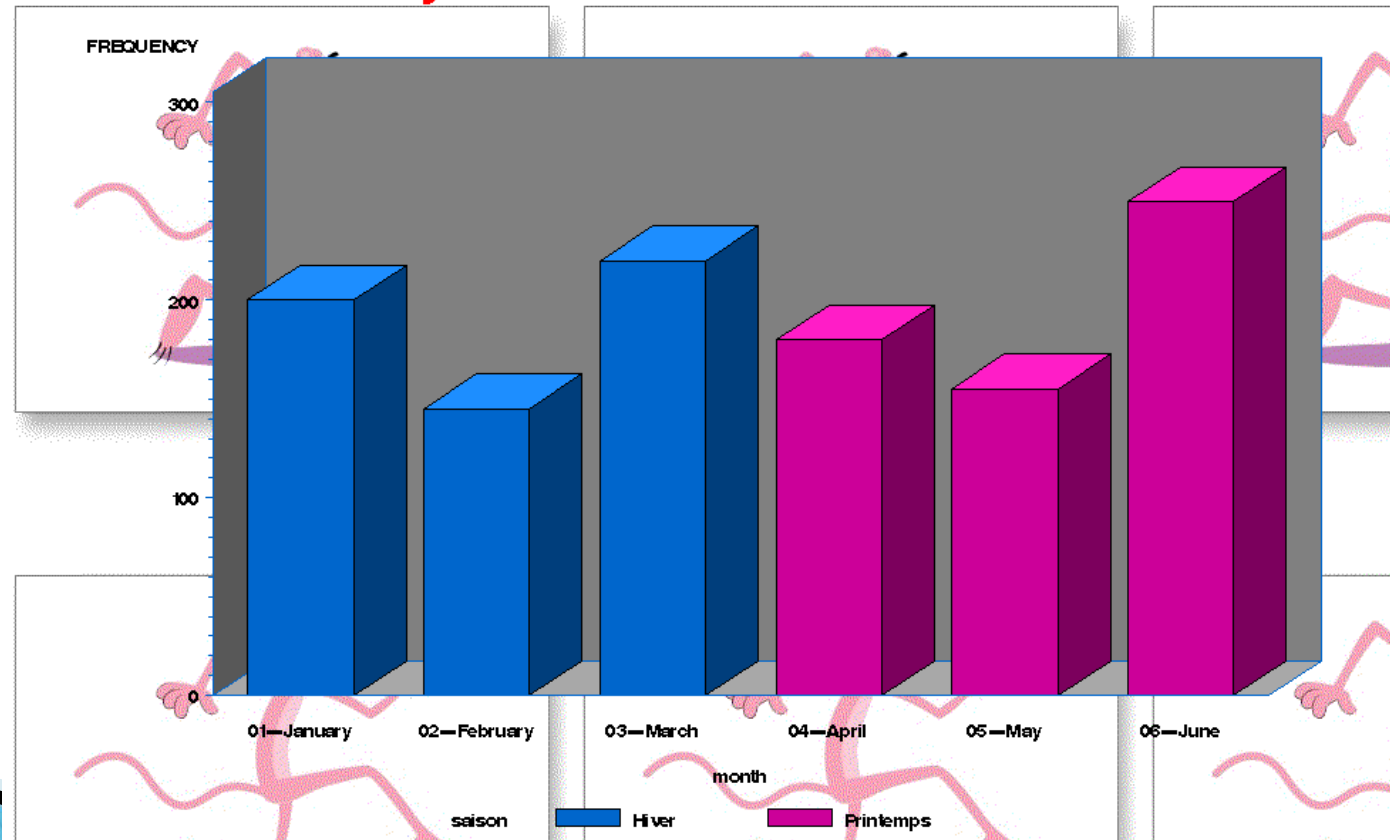


Graphiques sous SAS



Aujourd'hui, les graphiques avec SAS, c'est ça :

Projected Automobile Sales





Graphiques sous SAS

Mais pour cela, il faut parler très gentiment à SAS, avec des mots doux pleins d'options ...

```
goptions reset=all ctitle=cxFF0000 ftitle=swissb
          ctext=black htext=0.85 htitle=2.5 ftext=swissb
-         colors=(cx0066CC cxCC0099 )
          iback='FormationSAS/panthererose.jpg';
title 'Projected Automobile Sales';
data sales;
  length month $ 11;
  length saison $ 9;
  input month amount saison ;
  datalines;
01-January      200   Hiver
02-February     145   Hiver
03-March        220   Hiver
04-April        180   Printemps
05-May          155   Printemps
06-June         250   Printemps
;
proc gchart;
  vbar3d month / subgroup=saison freq=amount
            noheading coutline=black;
run;
```



Graphiques sous SAS

Mais pour cela, il faut parler très gentiment à SAS, avec des mots doux pleins d'options ...

Oublie tout ce qui
était défini avant

Couleur du titre (noir)

Fonte du titre :
Swiss, Gras

Couleur du texte
(noir)

```
goptions reset=all ctitle=cxFF0000 ftitle=swissb  
          ctext=black htext=0.85 htitle=2.5 ftext=swissb  
-         colors=(cx0066CC cxCC0099  
          iback='FormationSAS/panthererose.jpg';  
          title 'Projected Automobile Sales';
```

Fonte du texte :
Swiss, Gras

Image dans le fond

Hauteur du texte=0.85
Hauteur du titre=2.5

Titre du graphique

Couleur des différentes barres



Graphiques sous SAS

Mais pour cela, il faut parler très gentiment à SAS, avec des mots doux pleins d'options ...

```
data sales;
  length month $ 11;
  length saison $ 9;
  input month amount saison ;
  datalines;
01-January    200   Hiver
02-February   145   Hiver
03-March      220   Hiver
04-April      180   Printemps
05-May        155   Printemps
06-June       250   Printemps
;
proc gchart;
  vbar3d month / subgroup=saison freq=amount
  noheading coutline=black;
run;
```

Sous-groupe : saison

Nom de la procédure

Barres verticales

Pas de titre par défaut
(on en a déjà mis un par ailleurs)

La hauteur des barres
est indiquée par la
variable « amount »

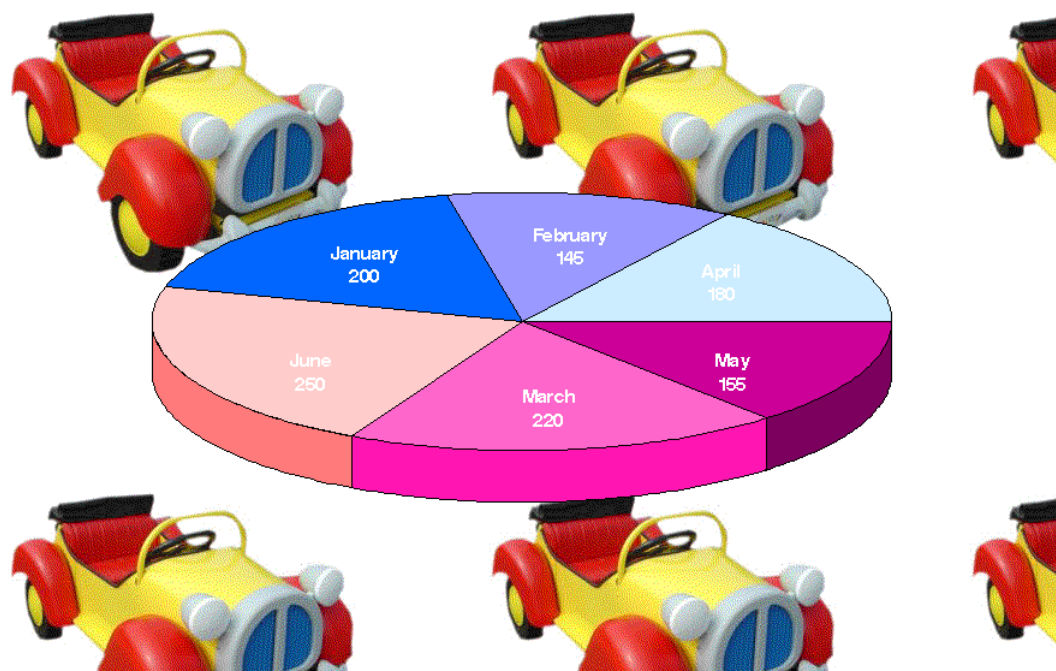
Couleur du contour
du graphique (noir)



Graphiques sous SAS

```
goptions reset=all ctitle=cx0000FF ftitle=swissb
          ctext=white htext=1.00 htitle=2.5 ftext=swissb
          colors=(cxCCECFF cx9999FF cx0066FF
                 cxFFCCCC cxFF66CC cxCC0099)
          iback='FormationSAS/voiture-ouioui.jpg';
title 'Projected Automobile Sales';
data sales;
  length month $ 9;
  input month amount;
  datalines;
January    200
February   145
March      220
April      180
May        155
June       250
;
proc gchart;
  pie3d month / freq=amount value=inside
  noheading coutline=black;
run;
quit;
```

Projected Automobile Sales





Graphiques sous SAS

Mais pour cela, il faut parler très gentiment à SAS, avec des mots doux pleins d'options ...

```
goptions reset=all ctitle=cxFF0000 ftitle=swissb
          ctext=black htext=0.85 htitle=2.5 ftext=swissb
          colors=(cx0066CC cxCC0099 cxFFFF00
                  cx00FF00 cxFF0000 cx00FFFF)
          iback='FormationSAS/panthererose.jpg';
title 'Projected Automobile Sales';
data sales;
  length month $ 11;
  length saison $ 9;
  input month amount saison ;
  datalines;
01-January    200  Hiver
02-February   145  Hiver
03-March      220  Hiver
04-April      180  Printemps
05-May        155  Printemps
06-June       250  Printemps
;
proc gchart;
  vbar3d month / subgroup=saison freq=amount
             noheading coutline=black;
run;
```



Graphiques sous SAS

Mais pour cela, il faut parler très gentiment à SAS, avec des mots doux pleins d'options ...

L'option « SYMBOL » peut aussi être utilisée pour définir la forme des courbes :

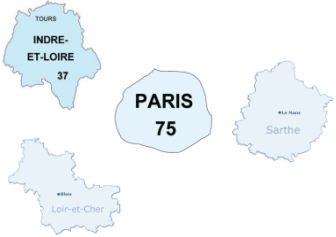
```
SYMBOL1    v = diamond    c=blue    i=join ;  
SYMBOL2    v = star       c=purple i=join;  
SYMBOL3    v = dot        c=red     ;
```

Donne aux courbes 1, 2 et 3 des symboles différents (losange, astérisque et point), des couleurs différentes (bleu, violet, rouge), et joint les points des lignes 1 et 2 par un trait continue

FONTES :

<http://support.sas.com/documentation/cdl/en/graphref/63022/HTML/default/viewer.htm#font-font-lists.htm>

COULEURS : http://thesasreference.wordpress.com/2008/11/21/couleurs_sas/



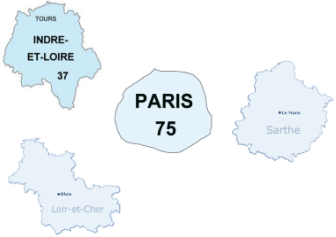
Exercice 4

Lire le Fichier « DonneesPopulation2.csv » qui contient :

- Année
- Population de l'Indre et Loire
- Population du Loir-et-Cher
- Population de la Sarthe
- Population de Paris

Réaliser le graphique de l'évolution de la population dans chaque département avec les années,

- Tous départements confondus sur le même graphique
- Définir un symbole différent pour chaque département
- Ajouter une ligne continue pour relier les points de la courbe
 - Tous sur la même échelle
 - Paris sur une échelle différente

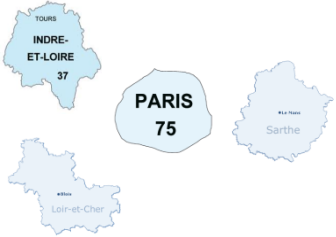


Exercice 4

```
data population;
infile "FormationSAS/DonneesPopulation2.csv" firstobs=2 dlm=';';
input annee Dep37 Dep41 Dep72 Dep75;

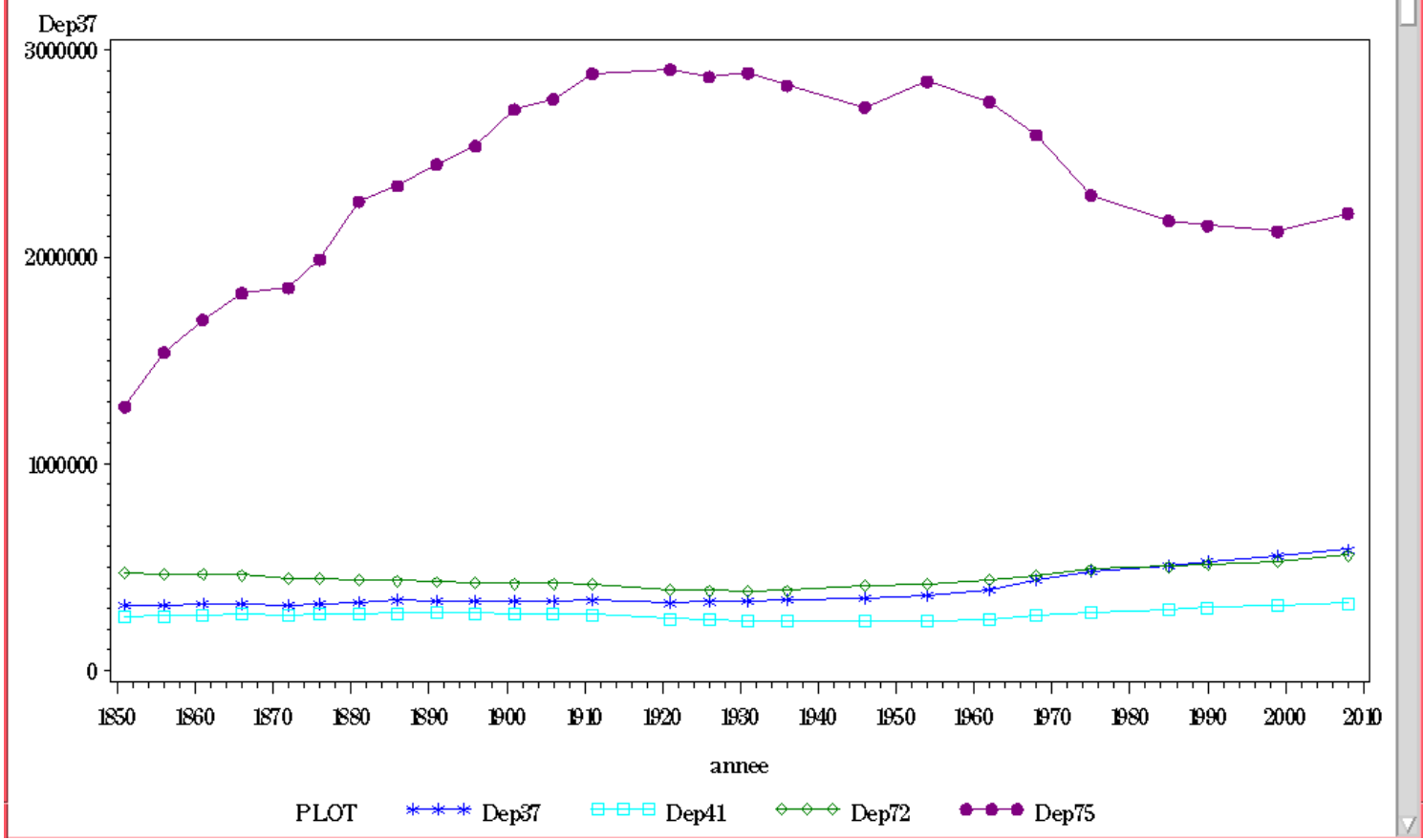
goptions reset=all ctitle=cxDDDDDD ftitle=CENTBE htitle=2.0 ctext=black ftext =CENTB htext
=1.0 ;
title 'Evolution de la population dans les principaux departements francais';

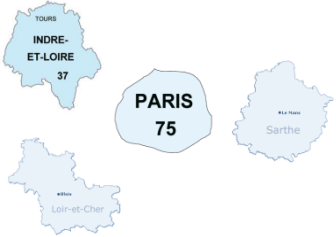
proc gplot;
plot (Dep37 Dep41 Dep72 Dep75) * annee / Overlay legend;
SYMBOL1 v=star      i=join c=blue;
SYMBOL2 v=square    i=join c=cyan;
SYMBOL3 v=diamond   i=join c=green;
SYMBOL4 v=dot       i=join c=purple;
run;
quit;
run;
```

Exercice 4

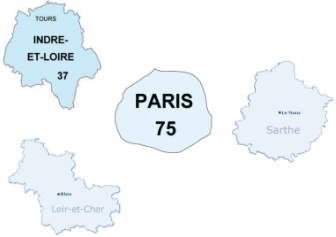
Evolution de la population dans les principaux departements francais



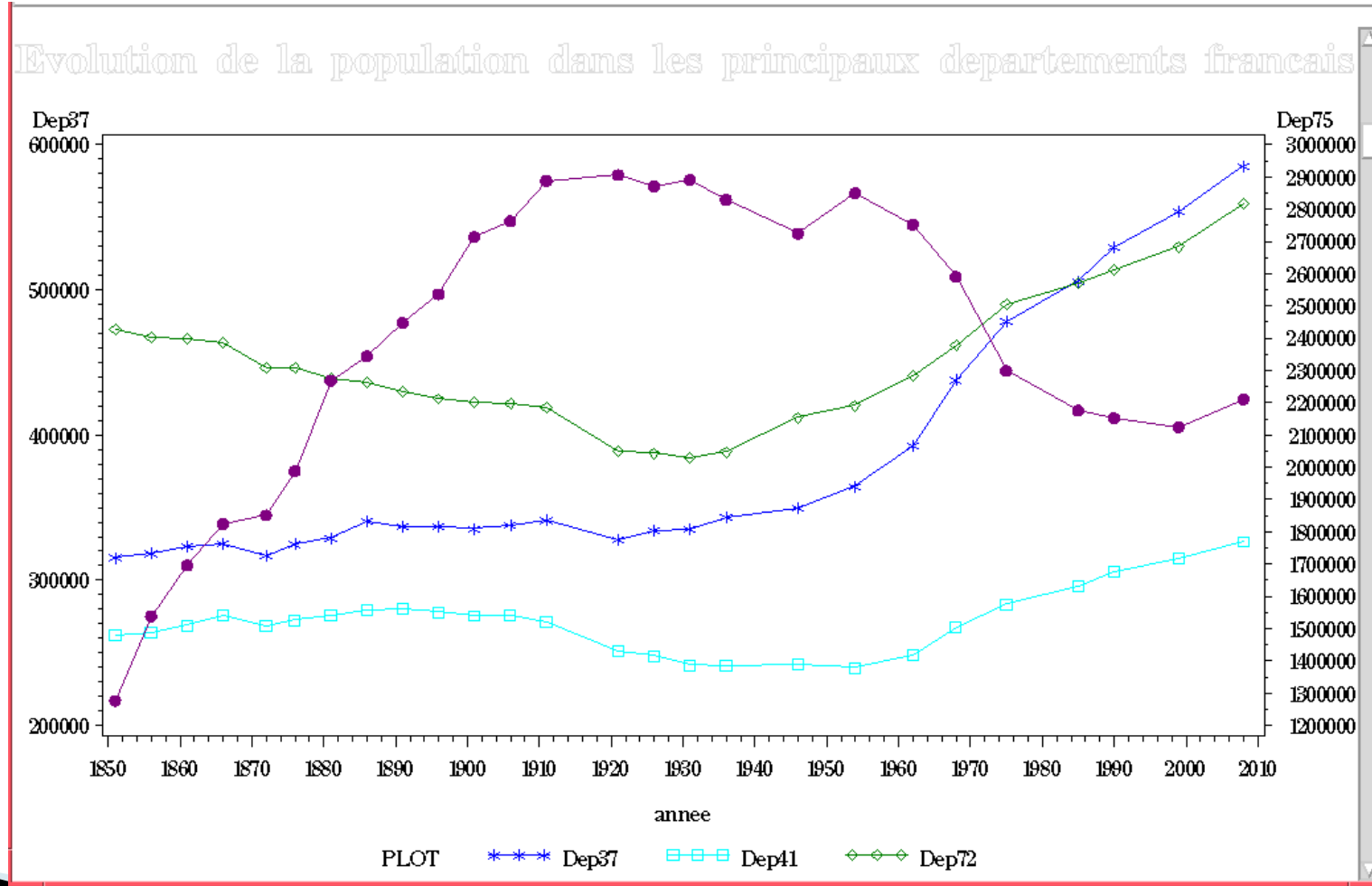


Exercice 4

```
proc gplot;  
  plot (Dep37 Dep41 Dep72) * annee / Overlay legend;  
  plot2 (Dep75) * annee / legend;  
  SYMBOL1 v=star i=join c=blue;  
  SYMBOL2 v=square i=join c=cyan;  
  SYMBOL3 v=diamond i=join c=green;  
  SYMBOL4 v=dot i=join c=purple;  
run;  
quit;  
run;
```



Exercice 4





Analyse de variance à plusieurs facteurs, sans interaction: proc GLM

```
SAS: Program Editor-Untitled
File Edit View Tools Run Solutions Help
NOTE: 9 Lines submitted.
00001 data seveso;
00002 infile "FormationSAS/DonneesSEVESO.csv" firstobs=2 dlm=';';
00003 input departement :$23. autorisation sevesoHR sevesoRM region :$13.
00004         pop2009 superficie taille :$10. densite densiteb :$19.;
00005 proc glm;
00006 class region taille densiteb;
00007 model autorisation sevesoHR sevesoRM = region taille densiteb;
00008 lsmeans densiteb / stderr pdiff=all;
00009 run;
00010
```

- ▶ **Class** : indique les variables qualitatives
- ▶ **Model** : précise le modèle d'analyse
 - Si plusieurs modèles différents selon la variable à analyser, il faut plusieurs instructions « proc glm »
- ▶ **Lsmeans** : donne la moyenne des moindres carrés de l'effet demandé
 - Stderr : donne les écarts-types des moyennes des moindres carrés
 - Pdiff=all : donne la significativité des différences entre tous les niveaux de l'effet demandé dans lsmeans



Analyse de variance à plusieurs facteurs, sans interaction: proc GLM

SAS: Output-Untitled-PROC GLM running

File Edit View Tools Solutions Help

The SAS System 11:12 Monday, March 28, 2011 121

The GLM Procedure

Dependent Variable: autorisation

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	10	7645072.01	764507.20	5.12	<.0001
Error	89	13294286.90	149374.01		
Corrected Total	99	20939358.91			

R-Square	Coeff Var	Root MSE	autorisation Mean
0.365105	75.70355	386.4893	510.5300

R² du modèle

Moyenne et coefficient de variation de la variable

Sommes de carrés du modèle



Analyse de variance à plusieurs facteurs, sans interaction: proc GLM

SAS: Output-Untitled-PROC GLM running

File Edit View Tools Solutions Help

The SAS System 11:12 Monday, March 28, 2011 122

The GLM Procedure

Dependent Variable: autorisation

Source	DF	Type I SS	Mean Square	F Value	Pr > F
region	4	4780579.240	1195144.810	8.00	<.0001
taille	3	502196.438	167398.813	1.12	0.3451
densiteb	3	2362296.333	787432.111	5.27	0.0022

Source	DF	Type III SS	Mean Square	F Value	Pr > F
region	4	2182435.485	545608.871	3.65	0.0084
taille	3	1103611.307	367870.436	2.46	0.0677
densiteb	3	2362296.333	787432.111	5.27	0.0022

• Les effets « région » et « densiteb » sont significatifs
L'effet taille est non significatif

• Les Type I et Type III donnent les mêmes résultats si les effectifs sont équilibrés dans les différents groupes.

Dans le cas contraire, c'est le type III qui donne les bons résultats



Analyse de variance à plusieurs facteurs, sans interaction: proc GLM

SAS: Output-Untitled-PROC GLM running

File Edit View Tools Solutions Help

The SAS System 11:12 Monday, March 28, 2011 127

The GLM Procedure
Least Squares Means
Adjustment for Multiple Comparisons: Tukey-Kramer

densiteb	autorisation LSMEAN	Standard Error	Pr > t	LSMEAN Number	
Densité Faible	470.455092	79.403853	<.0001	1	ab
Densité Moyenne	716.934627	81.743738	<.0001	2	a
Densité très fai	289.662469	87.993459	0.0014	3	b
Forte Densité	731.329159	138.822211	<.0001	4	ab

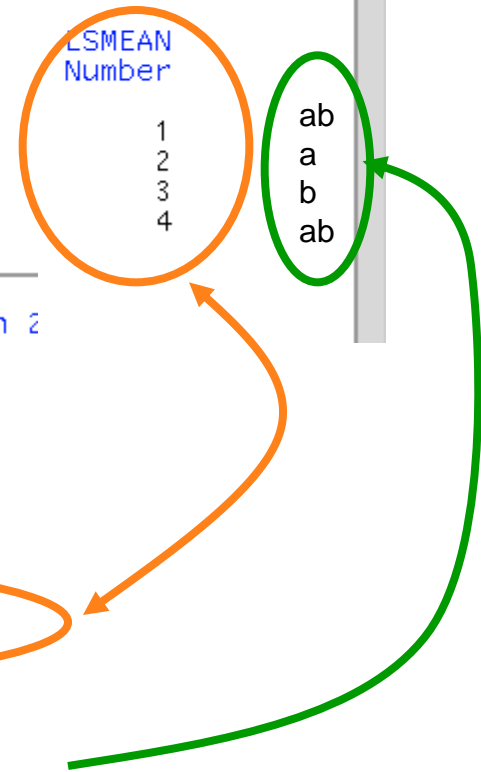
The SAS System 11:12 Monday, March 28, 2011 127

The GLM Procedure
Least Squares Means
Adjustment for Multiple Comparisons: Tukey-Kramer

Least Squares Means for effect densiteb
Pr > |t| for H0: LSmean(i)=LSmean(j)

Dependent Variable: autorisation

i/j	1	2	3	4
1		0.0864	0.3169	0.4654
2	0.0864		0.0014	0.9998
3	0.3169	0.0014		0.0776
4	0.4654	0.9998	0.0776	





Analyse de variance à plusieurs facteurs, sans interaction: proc GLM

Si l'effet n'est pas listé dans l'instruction « class », il est traité comme une variable continue

```
SAS: Program Editor--seveso.sas
File Edit View Tools Run Solutions Help
NOTE: 8 Line(s) recalled.
00001 data seveso;
00002 infile "FormationSAS/DonneesSEVES0.csv" firstobs=2 dlm=';';
00003 input departement :$23. autorisation sevesoHR sevesoRM region :$13.
00004         pop2009 superficie taille :$10. densite densiteb :$19.;
00005 proc glm;
00006 class region taille densiteb;
00007 model sevesoHR = region taille densite / e3;
00008 run;
```

L'option e3 fait que l'on ne sortira que les sommes de carrés de type III



Analyse de variance à plusieurs facteurs, sans interaction: proc GLM

Si l'effet n'est pas listé dans l'instruction « class », il est traité comme une variable continue

The GLM Procedure

Dependent Variable: sevesoHR

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	8	335.663234	41.957904	0.74	0.6555
Error	91	5156.096766	56.660404		
Corrected Total	99	5491.760000			

R-Square	Coeff Var	Root MSE	sevesoHR Mean
0.061121	119.1030	7.527311	6.320000

Source	DF	Type III SS	Mean Square	F Value	Pr > F
region	4	88.2508403	22.0627101	0.39	0.8157
taille	3	216.8101164	72.2700388	1.28	0.2876
densite	1	21.1767900	21.1767900	0.37	0.5425



Les Type I sont partis !



Analyse de variance à plusieurs facteurs avec interaction: proc GLM

- ▶ Dans l'instruction **model** :
 - `model ma_variable = effet1 effet2;`
⇒ modèle sans interaction
 - `model ma_variable = effet1 * effet2;`
⇒ modèle avec uniquement l'interaction entre effet1 et effet 2
 - `model ma_variable = effet1 effet2 effet1 * effet2;`
⇒ modèle avec les 2 effets simples plus l'interaction
 - `model ma_variable = effet1 | effet2;`
⇒ modèle avec les 2 effets simples plus l'interaction



Analyse de variance à plusieurs facteurs avec interaction: proc GLM

- ▶ Dans l'exemple précédent :

```
SAS: Program Editor--seveso.sas--PROC GLM running
File Edit View Tools Run Solutions Help
NOTE: 9 Line(s) recalled.
00001 data seveso;
00002 infile "FormationSAS/DonneesSEVESO.csv" firstobs=2 dlm=';';
00003 input departement :$23. autorisation sevesoHR sevesoRM region :$13.
00004         pop2009 superficie taille :$10. densite densiteb :$19.;
00005 proc glm;
00006 class region taille densiteb;
00007 model sevesoHR = region taille densiteb region*densiteb;
00008 lsmeans region*densiteb / stderr pdiff=all;
00009 run;
00010
00011
00012
```

- ▶ On a ici ajouté l'interaction entre « region » et « densiteb » au modèle



Analyse de variance à plusieurs facteurs avec interaction: proc GLM

the GLM procedure

Dependent Variable: sevesoHR

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	20	2695.792690	134.789635	3.81	<.0001
Error	79	2795.967310	35.391991		
Corrected Total	99	5491.760000			

R-Square	Coeff Var	Root MSE	sevesoHR Mean
0.490880	94.13160	5.949117	6.320000

Source	DF	Type I SS	Mean Square	F Value	Pr > F
region	4	91.098913	22.774728	0.64	0.6331
taille	3	223.387531	74.462510	2.10	0.1064
densiteb	3	1538.433256	512.811085	14.49	<.0001
region*densiteb	10	842.872990	84.287299	2.38	0.0160

Source	DF	Type III SS	Mean Square	F Value	Pr > F
region	4	340.147645	85.036911	2.40	0.0567
taille	3	658.414040	219.471347	6.20	0.0008
densiteb	3	1203.389400	401.129800	11.33	<.0001
region*densiteb	10	842.872990	84.287299	2.38	0.0160



Analyse de variance à plusieurs facteurs avec interaction: proc GLM

The GLM Procedure
Least Squares Means
Adjustment for Multiple Comparisons: Tukey-Kramer

region	densiteb	sevesoHR LSMEAN	Standard Error	Pr > t	LSMEAN Number
Ile de France	Densité Moyenne	16.1478799	6.0484030	0.0092	1
Ile de France	Forte Densité	7.7536404	2.6433919	0.0044	2
Nord Est	Densité Faible	2.5114281	2.0224885	0.2180	3
Nord Est	Densité Moyenne	12.5681284	2.1602171	<.0001	4
Nord Est	Densité très fai	0.4459365	2.7380558	0.8710	5
Nord Est	Forte Densité	29.1478799	6.0484030	<.0001	6
Nord Ouest	Densité Faible	4.8385271	2.0571693	0.0212	7
Nord Ouest	Densité Moyenne	9.1324702	2.3823183	0.0003	8
Nord Ouest	Densité très fai	0.7443328	3.5267404	0.8334	9
Nord Ouest	Forte Densité	6.4679261	6.1092973	0.2930	10
Sud Est	Densité Faible	5.0459365	2.7380558	0.0691	11
Sud Est	Densité Moyenne	8.4110598	2.0376308	<.0001	12
Sud Est	Densité très fai	3.6539515	2.1386105	0.0915	13
Sud Est	Forte Densité	25.6945774	3.5533524	<.0001	14
Sud Ouest	Densité Faible	6.2908250	1.9232627	0.0016	15
Sud Ouest	Densité Moyenne	8.6173734	3.4539375	0.0147	16
Sud Ouest	Densité très fai	2.5446587	2.0284376	0.2134	17
Sud Ouest	Forte Densité	7.4679261	6.1092973	0.2252	18



Analyse de variance à plusieurs facteurs avec interaction: proc GLM

Least Squares Means for effect region*densiteb
Pr > |t| for H0: LSMean(i)=LSMean(j)

Dependent Variable: sevesoHR

i/j	1	2	3	4	5	6	7	8	9
1		0.9983	0.7884	1.0000	0.6361	0.9841	0.9427	0.9998	0.7577
2	0.9983		0.9865	0.9909	0.9253	0.1394	1.0000	1.0000	0.9855
3	0.7884	0.9865		0.1056	1.0000	0.0089	1.0000	0.7912	1.0000
4	1.0000	0.9909	0.1056		0.0766	0.4682	0.4587	0.9998	0.3206
5	0.6361	0.9253	1.0000	0.0766		0.0053	0.9965	0.5446	1.0000
6	0.9841	0.1394	0.0089	0.4682	0.0053		0.0267	0.1933	0.0140
7	0.9427	1.0000	1.0000	0.4587	0.9965	0.0267		0.9913	0.9999
8	0.9998	1.0000	0.7912	0.9998	0.5446	0.1933	0.9913		0.8427
9	0.7577	0.9855	1.0000	0.3206	1.0000	0.0140	0.9999	0.8427	
10	0.9996	1.0000	1.0000	1.0000	1.0000	0.4484	1.0000	1.0000	1.0000
11	0.9666	1.0000	1.0000	0.7836	0.9988	0.0471	1.0000	0.9993	0.9999
12	0.9989	1.0000	0.8516	0.9928	0.6884	0.1247	0.9990	1.0000	0.9225
13	0.8797	0.9985	1.0000	0.2325	1.0000	0.0156	1.0000	0.9650	1.0000
14	0.9957	0.0054	<.0001	0.1364	<.0001	1.0000	0.0005	0.0309	0.0008
15	0.9815	1.0000	0.9965	0.7430	0.9509	0.0456	1.0000	1.0000	0.9956
16	0.9998	1.0000	0.9858	0.9999	0.9193	0.2639	1.0000	1.0000	0.9773
17	0.7868	0.9870	1.0000	0.1069	1.0000	0.0086	1.0000	0.8127	1.0000
18	0.9999	1.0000	1.0000	1.0000	0.9999	0.5306	1.0000	1.0000	1.0000

Bonne chance pour les petites lettres ...





Analyse de variance à plusieurs facteurs imbriqués: proc GLM

- ▶ Pour que l'effet 2 soit imbriqué dans l'effet 1, on écrit l'instruction model de la proc GLM comme suit :
- ▶ **model ma_variable = effet2(effet1);**

```
SAS: Program Editor--seveso.sas--PROC GLM running
File Edit View Tools Run Solutions Help
00001 data seveso;
00002 infile "FormationSAS/DonneesSEVESO.csv" firstobs=2 dlm=';';
00003 input departement :$23. autorisation sevesoHR sevesoRM region :$13.
00004         pop2009 superficie taille :$10. densite densiteb :$19.;
00005 proc glm;
00006 class region taille densiteb;
00007 model sevesoHR = region taille densiteb(region);
00008 lsmeans densiteb(region) / stderr pdiff=all;
00009 run;
00010
00011
00012
```



Analyse de variance à plusieurs facteurs imbriqués: proc GLM

THE GLM PROCEDURE

Dependent Variable: sevesoHR

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	20	2695.792690	134.789635	3.81	<.0001
Error	79	2795.967310	35.391991		
Corrected Total	99	5491.760000			

R-Square Coeff Var Root MSE sevesoHR Mean
 0.490880 94.13160 5.949117 6.320000

Source	DF	Type I SS	Mean Square	F Value	Pr > F
region	4	91.098913	22.774728	0.64	0.6331
taille	3	223.387531	74.462510	2.10	0.1064
region(densiteb)	13	2381.306246	183.177404	5.18	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
region	4	345.395710	86.348927	2.44	0.0537
taille	3	658.414040	219.471347	6.20	0.0008
region(densiteb)	13	2381.306246	183.177404	5.18	<.0001



Analyse de variance à plusieurs facteurs imbriqués: proc GLM

SAS: Output–Untitled–PROC GLM running

File Edit View Tools Solutions Help

The SAS System

14:26 Wednesday, March 30, 2011

The GLM Procedure
Least Squares Means

Adjustment for Multiple Comparisons: Tukey–Kramer

region	densiteb	sevesoHR LSMEAN	Standard Error	Pr > t	LSMEAN Number
Nord Est	Densité Faible	2.5114281	2.0224885	0.2180	1
Nord Ouest	Densité Faible	4.8385271	2.0571693	0.0212	2
Sud Est	Densité Faible	5.0459365	2.7380558	0.0691	3
Sud Ouest	Densité Faible	6.2908250	1.9232627	0.0016	4
Ile de France	Densité Moyenne	16.1478799	6.0484030	0.0092	5
Nord Est	Densité Moyenne	12.5681284	2.1602171	<.0001	6
Nord Ouest	Densité Moyenne	9.1324702	2.3823183	0.0003	7
Sud Est	Densité Moyenne	8.4110598	2.0376308	<.0001	8
Sud Ouest	Densité Moyenne	8.6173734	3.4539375	0.0147	9
Nord Est	Densité très fai	0.4459365	2.7380558	0.8710	10
Nord Ouest	Densité très fai	0.7443328	3.5267404	0.8334	11
Sud Est	Densité très fai	3.6539515	2.1386105	0.0915	12
Sud Ouest	Densité très fai	2.5446587	2.0284376	0.2134	13
Ile de France	Forte Densité	7.7536404	2.6433919	0.0044	14
Nord Est	Forte Densité	29.1478799	6.0484030	<.0001	15
Nord Ouest	Forte Densité	6.4679261	6.1092973	0.2930	16
Sud Est	Forte Densité	25.6945774	3.5533524	<.0001	17
Sud Ouest	Forte Densité	7.4679261	6.1092973	0.2252	18



Exercice 2



Reprendre le fichier « Céréales »

Réaliser une analyse de variance à effets fixes simples, sans interaction

Variables :

- calorie
- lipides
- glucides

Effets :

- cat_cereale
- parfum
- cat_prix
- cer_complete
- Vit_Fer

-Demander les moyennes des moindres carrés avec l'option lsmeans pour l'effet catégorie de céréales et parfum.

-Demander toutes les différences significatives

-Demander à sauvegarder dans un fichier séparé les valeurs prédites et les résidus de l'analyse (pour vérifier si les conditions de l'ANOVA sont réunies)



Exercice 2



```
proc glm data=cerealesmerge;

class cat_cereale parfum cat_prix cer_complete Vit_Fer ;
model calorie lipides glucides =
    cat_cereale parfum cat_prix cer_complete Vit_Fer;

lsmeans cat_cereale parfum cat_prix cer_complete Vit_Fer / pdiff = all ;

title 'Analyses de variances sur les calories, les lipides, les glucides';

output out=resglm p=calpred lippred glupred r=calres lipres glures;

proc gplot data=resglm;
plot calres*calpred lipres*lippred glures*glupred / legend ;
SYMBOL1 v=dot i=rl c=blue;
```



Exercice 2



Class Level Information		
Class	Levels	Values
cat_cereale	4	Corn Flakes Fourree Muesli Soufflee
parfum	4	Chocolat Fruits Nature Sucre
cat_prix	3	Distributeur Marque Premier Prix
cer_complete	2	non oui
Vit_Fer	2	non oui

Number of observations	113
------------------------	-----

Dependent Variables With Equivalent Missing Value Patterns		
Pattern	Obs	Dependent Variables
1	113	calorie lipides
2	112	glucides

NOTE: Variables in each group are consistent with respect to the presence or absence of missing values.



Exercice 2

Calories

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	10	35314.4347	3531.4435	4.58	<.0001
Error	102	78726.0078	771.8236		
Corrected Total	112	114040.4425			

R-Square	Coeff Var	Root MSE	calorie Mean
0.309666	7.139068	27.78171	389.1504

Source	DF	Type I SS	Mean Square	F Value	Pr > F
cat_cereale	3	24340.65285	8113.55095	10.51	<.0001
parfum	3	6976.38800	2325.46267	3.01	0.0335
cat_prix	2	422.02072	211.01036	0.27	0.7614
cer_complete	1	3336.13246	3336.13246	4.32	0.0401
Vit_Fer	1	239.24067	239.24067	0.31	0.5789

Source	DF	Type III SS	Mean Square	F Value	Pr > F
cat_cereale	3	12100.79378	4033.59793	5.23	0.0021
parfum	3	5966.21649	1988.73883	2.58	0.0579
cat_prix	2	368.70756	184.35378	0.24	0.7880
cer_complete	1	3000.19485	3000.19485	3.89	0.0514
Vit_Fer	1	239.24067	239.24067	0.31	0.5789



Exercice 2



Lipides

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	10	2381.039662	238.103966	18.20	<.0001
Error	102	1334.259984	13.080980		
Corrected Total	112	3715.299646			

R-Square	Coeff Var	Root MSE	lipides Mean
0.640874	57.39283	3.616764	6.301770

Source	DF	Type I SS	Mean Square	F Value	Pr > F
cat_cereale	3	2035.339984	678.446661	51.87	<.0001
parfum	3	273.824617	91.274872	6.98	0.0003
cat_prix	2	41.067932	20.533966	1.57	0.2131
cer_complete	1	26.256026	26.256026	2.01	0.1596
Vit_Fer	1	4.551103	4.551103	0.35	0.5566

Source	DF	Type III SS	Mean Square	F Value	Pr > F
cat_cereale	3	1644.818218	548.272739	41.91	<.0001
parfum	3	260.461826	86.820609	6.64	0.0004
cat_prix	2	54.520194	27.260097	2.08	0.1297
cer_complete	1	29.016960	29.016960	2.22	0.1395
Vit_Fer	1	4.551103	4.551103	0.35	0.5566



Exercice 2



cat_cereale	calorie LSMEAN	LSMEAN Number
Corn Flakes	378.223535	1
Fourree	414.055390	2
Muesli	391.760562	3
Soufflee	380.888176	4

cat_cereale	lipides LSMEAN	LSMEAN Number
Corn Flakes	3.3600511	1
Fourree	13.6012064	2
Muesli	10.5312366	3
Soufflee	3.1450149	4

Least Squares Means for effect cat_cereale Pr > t for H0: LSMEAN(i)=LSMEAN(j)				
Dependent Variable: calorie				
i/j	1	2	3	4
1		0.0099	0.4216	0.9898
2	0.0099		0.2325	0.0039
3	0.4216	0.2325		0.5697
4	0.9898	0.0039	0.5697	

Least Squares Means for effect cat_cereale Pr > t for H0: LSMEAN(i)=LSMEAN(j)				
Dependent Variable: lipides				
i/j	1	2	3	4
1		<.0001	<.0001	0.9975
2	<.0001		0.1893	<.0001
3	<.0001	0.1893		<.0001
4	0.9975	<.0001	<.0001	



Exercice 2



parfum	calorie LSMEAN	LSMEAN Number
Chocolat	399.862319	1
Fruits	392.160968	2
Nature	380.223548	3
Sucre	392.680828	4

parfum	lipides LSMEAN	LSMEAN Number
Chocolat	9.44138666	1
Fruits	8.66507138	2
Nature	5.66290577	3
Sucre	6.86814510	4

Least Squares Means for effect parfum Pr > t for H0: LSMean(i)=LSMean(j)				
Dependent Variable: calorie				
i/j	1	2	3	4
1		0.9031	0.0431	0.8049
2	0.9031		0.6217	1.0000
3	0.0431	0.6217		0.4683
4	0.8049	1.0000	0.4683	

Least Squares Means for effect parfum Pr > t for H0: LSMean(i)=LSMean(j)				
Dependent Variable: lipides				
i/j	1	2	3	4
1		0.9518	0.0008	0.0697
2	0.9518		0.0962	0.6598
3	0.0008	0.0962		0.7016
4	0.0697	0.6598	0.7016	



Exercice 2



Glucides

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	10	7408.107326	740.810733	35.30	<.0001
Error	101	2119.386960	20.984029		
Corrected Total	111	9527.494286			

R-Square	Coeff Var	Root MSE	glucides Mean
0.777550	6.070774	4.580833	75.45714

cat_cereale	glucides LSMEAN	LSMEAN Number
Corn Flakes	78.3390597	1
Fourree	64.7455143	2
Muesli	64.5286405	3
Soufflee	80.8458851	4

Source	DF	Type I SS	Mean Square	F Value	Pr > F
cat_cereale	3	6462.476291	2154.158764	102.66	<.0001
parfum	3	424.684809	141.561603	6.75	0.0003
cat_prix	2	380.981278	190.490639	9.08	0.0002
cer_complete	1	64.329423	64.329423	3.07	0.0830
Vit_Fer	1	75.635525	75.635525	3.60	0.0605

Source	DF	Type III SS	Mean Square	F Value	Pr > F
cat_cereale	3	5073.132935	1691.044312	80.59	<.0001
parfum	3	353.452472	117.817491	5.61	0.0013
cat_prix	2	412.742337	206.371169	9.83	0.0001
cer_complete	1	46.127271	46.127271	2.20	0.1413
Vit_Fer	1	75.635525	75.635525	3.60	0.0605

Least Squares Means for effect cat_cereale Pr > t for H0: LSMean(i)=LSMean(j)				
Dependent Variable: glucides				
i/j	1	2	3	4
1		<.0001	<.0001	0.2990
2	<.0001		0.9995	<.0001
3	<.0001	0.9995		<.0001
4	0.2990	<.0001	<.0001	



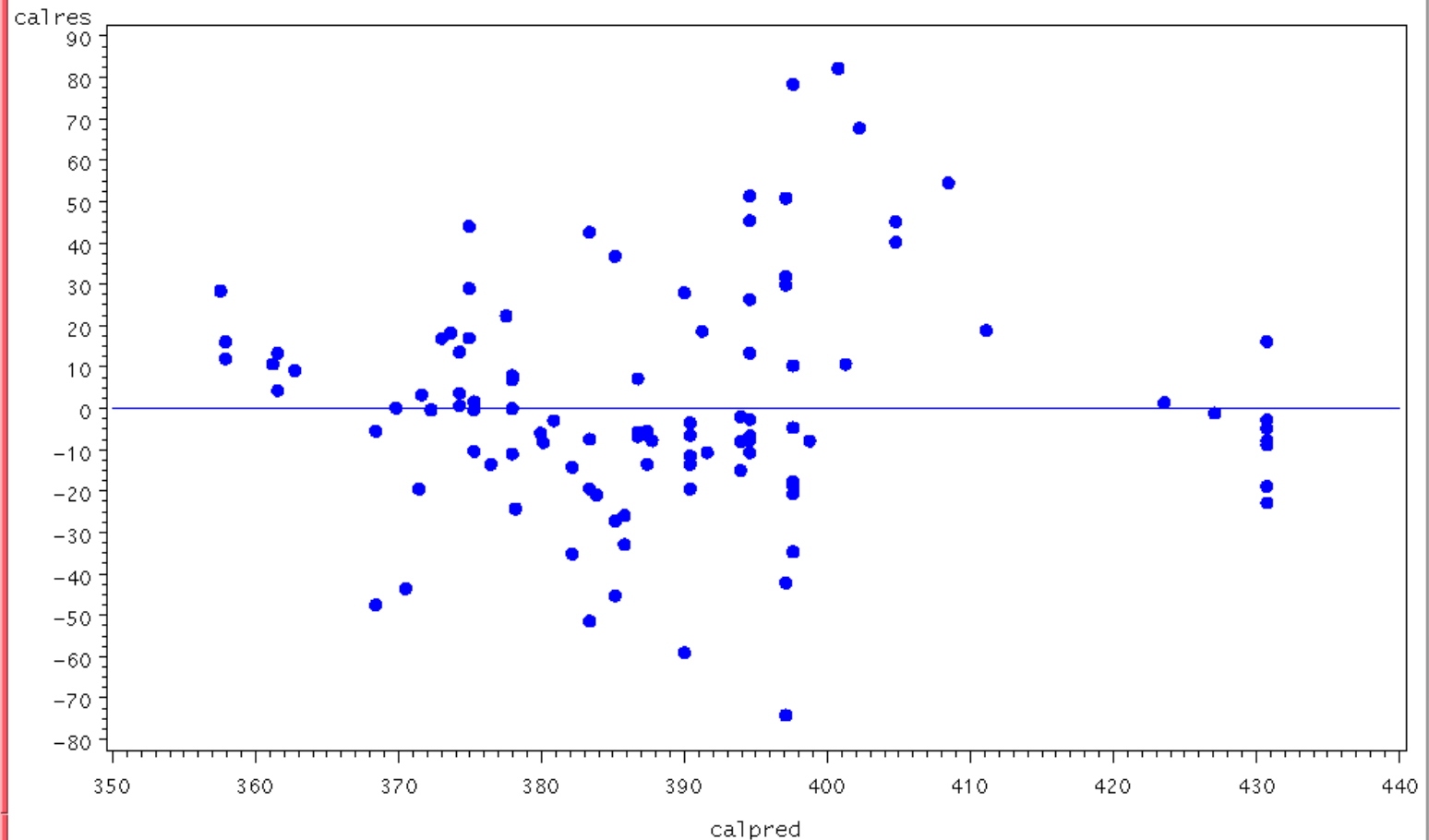
Exercice 2



SAS: GRAPH1 - WORK.GSEG.GPLOT9

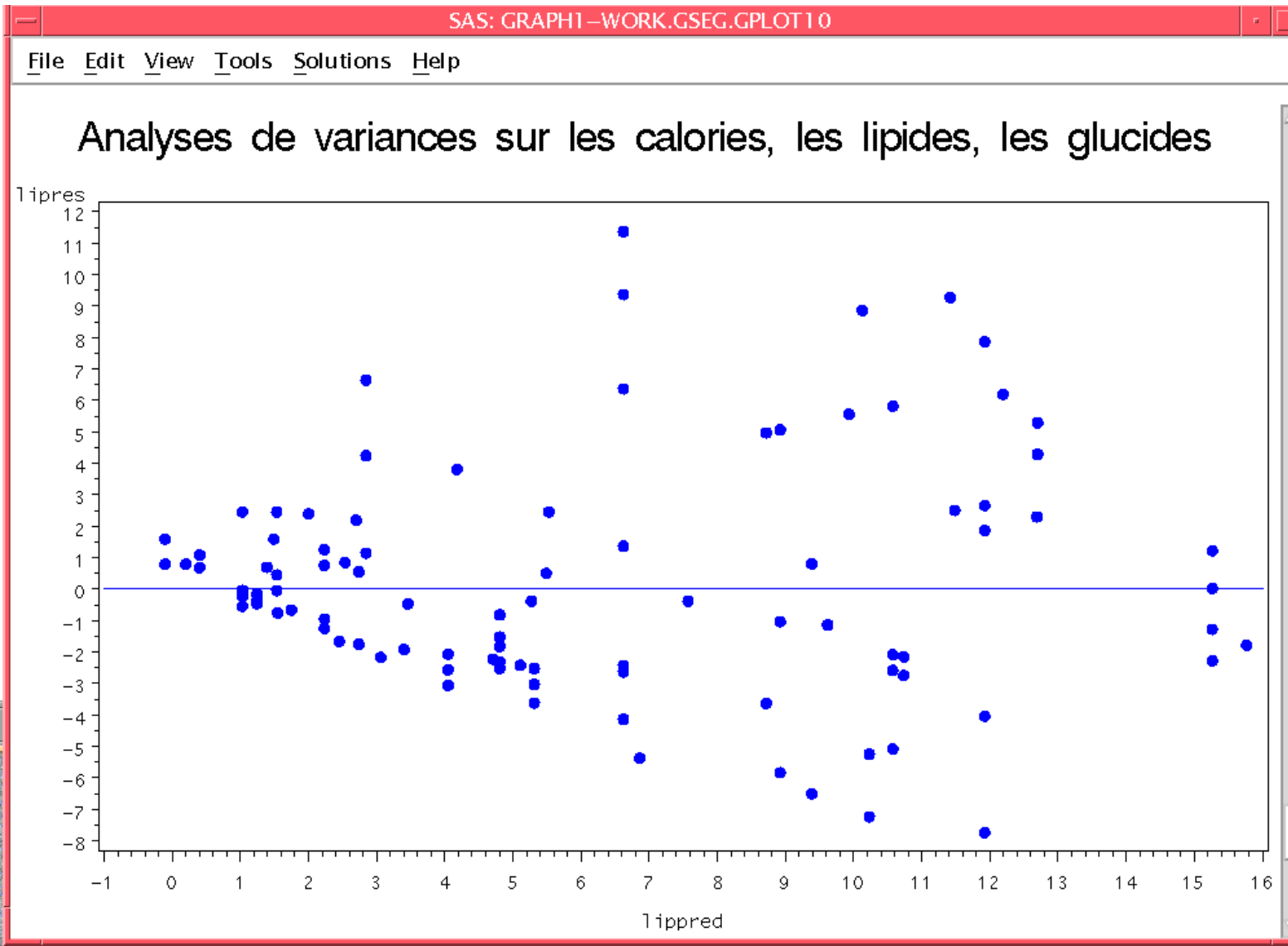
File Edit View Tools Solutions Help

Analyses de variances sur les calories, les lipides, les glucides





Exercice 2

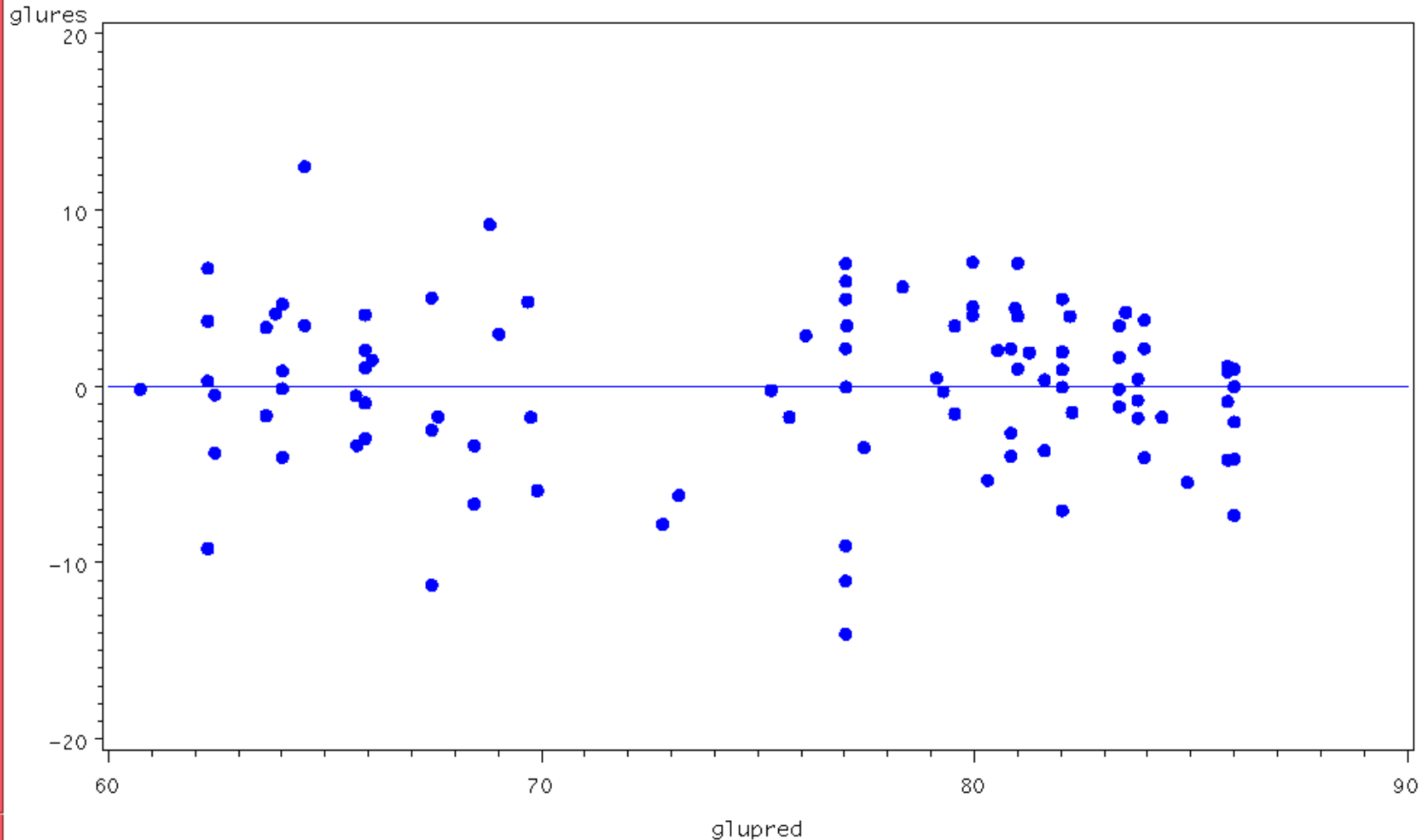




Exercice 2



Analyses de variances sur les calories, les lipides, les glucides





Analyse de variance avec effets aléatoires : proc mixed

```
1 data ronfle;  
2 infile "FormationSAS/DonneesRonflement.csv" firstobs=2 dlm=';';  
3 input patient $ age pere poids taille alcool sexe ronfle tabac;  
4
```





Analyse de variance avec effets aléatoires : proc mixed

```
proc mixed maxiter=1000 method=REML;  
  class sexe tabac pere;  
  id patient;  
  model ronfle = sexe tabac alcool /outp=sortie;  
  random pere;  
  estimate 'tabac vs sans tabac' tabac 1 -1;  
proc print data=sortie;  
run;
```

Variables à conserver dans le fichier de sortie

Spécifie les effets aléatoires (ici, le père)

Estime les contrastes entre différents niveaux d'effets (ici fumeur vs non fumeur)

Maxiter : nombre d'itérations maximal (par défaut : 50)

Method : méthode utilisée pour l'estimation (par défaut : REML)

Outp : précise que le fichier « sortie » contiendra les valeurs prédites par le modèle



Analyse de variance avec effets aléatoires : proc mixed

The Mixed Procedure

Iteration History

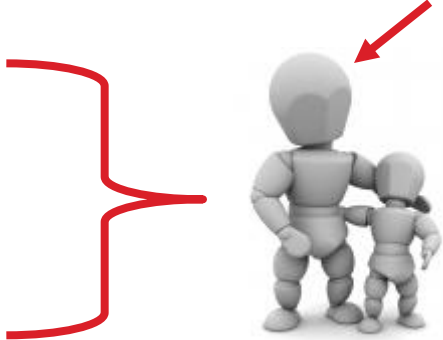
Iteration	Evaluations	-2 Res Log Like	Criterion
0	1	140.48748158	
1	2	140.44021082	0.00000000



Convergence criteria met.

Covariance Parameter Estimates

Cov Parm	Estimate
pere	0.002408
Residual	0.2090



Variance due au père et variance résiduelle



$$h^2 = \frac{4 \times \sigma_P^2}{\sigma_P^2 + \sigma_R^2}$$

Fit Statistics

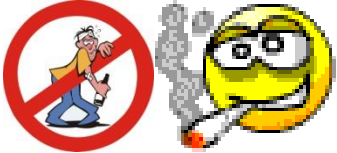
-2 Res Log Likelihood	140.4
AIC (smaller is better)	144.4
AICC (smaller is better)	144.6
BIC (smaller is better)	145.0



Analyse de variance avec effets aléatoires : proc mixed

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
sexe	1	87	1.51	0.2220
tabac	1	87	5.09	0.0266
alcool	1	87	5.27	0.0241

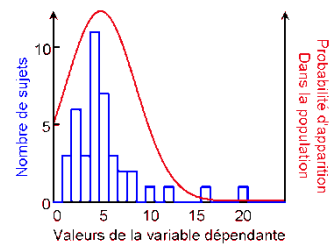


Estimates

Label	Estimate	Standard Error	DF	t Value	Pr > t
tabac vs sans tabac	0.2328	0.1032	87	2.26	0.0266

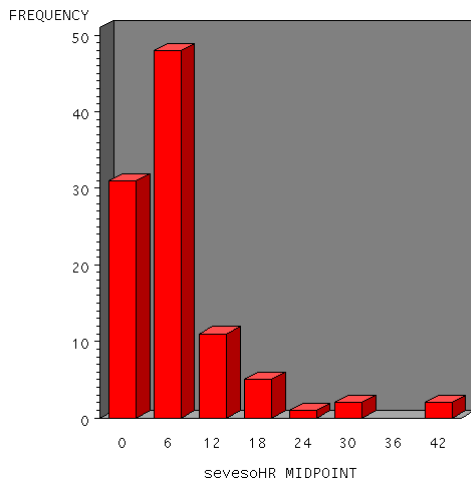


Fumer est associé avec une fréquence de ronflement plus faible

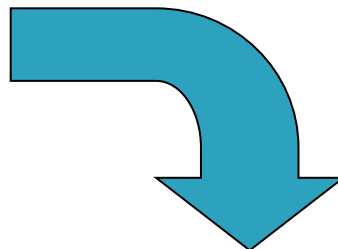


Statistiques non paramétriques : proc npar1way

Si les données ne sont pas normalement distribuées



Statistiques non paramétriques

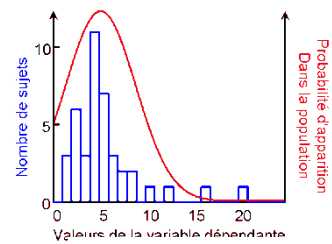


SAS: Program Editor—seveso.sas

File Edit View Tools Run Solutions Help

```
NOTE: 8 Lines submitted.
00010 data seveso;
00011 infile "FormationSAS/DonneesSEVESO.csv" firstobs=2 dlm=';';
00012 input departement :$23. autorisation sevesoHR sevesoRM region :$13.
00013         pop2009 superficie taille :$10. densite densiteb :$19.;
00014 proc npar1way;
00015 class region;
00016 var sevesoHR;
00017 run;
00018
```

1 seul effet possible



Statistiques non paramétriques : proc npar1way



Pour ne pas vous brusquer, SAS vous fait quand même une analyse variance ...

The SAS System

14:26 Wednesday, March 30,

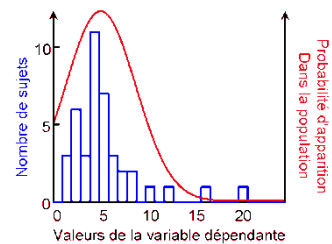
The NPAR1WAY Procedure

Analysis of Variance for Variable sevesoHR
Classified by Variable region

region	N	Mean
Ile de France	8	4.625000
Nord Est	23	6.913043
Sud Ouest	23	5.000000
Nord Ouest	21	7.000000
Sud Est	25	6.960000

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Among	4	91.098913	22.774728	0.4006	0.8078
Within	95	5400.661087	56.849064		

Average scores were used for ties.



Statistiques non paramétriques : proc npar1way



Après, on passe aux choses sérieuses ...

The NPAR1WAY Procedure

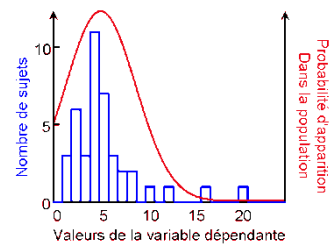
Van der Waerden Scores (Normal) for Variable sevesoHR
Classified by Variable region

region	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
Ile de France	8	-1.826809	0.0	2.597173	-0.228351
Nord Est	23	1.083799	0.0	4.028758	0.047122
Sud Ouest	23	-2.356162	0.0	4.028758	-0.102442
Nord Ouest	21	4.517209	0.0	3.899286	0.215105
Sud Est	25	-1.418037	0.0	4.145363	-0.056721

Average scores were used for ties.

Van der Waerden One-Way Analysis

Chi-Square	1.9222
DF	4
Pr > Chi-Square	0.7501



Statistiques non paramétriques : proc npar1way



Après, on passe aux choses sérieuses ...

The NPAR1WAY Procedure

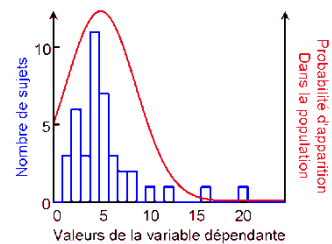
Wilcoxon Scores (Rank Sums) for Variable sevesoHR
Classified by Variable region

region	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
Ile de France	8	366.50	404.00	78.387239	45.812500
Nord Est	23	1197.50	1161.50	121.595006	52.065217
Sud Ouest	23	1090.50	1161.50	121.595006	47.413043
Nord Ouest	21	1198.50	1060.50	117.687314	57.071429
Sud Est	25	1197.00	1262.50	125.114342	47.880000

Average scores were used for ties.

Kruskal-Wallis Test

Chi-Square 1.8324
DF 4
Pr > Chi-Square 0.7666



Statistiques non paramétriques : proc npar1way



Après, on passe aux choses sérieuses ...

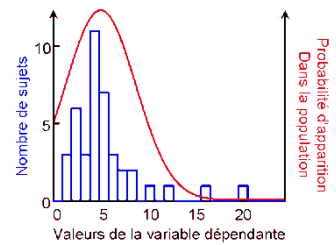
The NPAR1WAY Procedure

Kolmogorov-Smirnov Test for Variable sevesoHR
Classified by Variable region

region	N	EDF at Maximum	Deviation from Mean at Maximum
Ile de France	8	0.375000	-0.070711
Nord Est	23	0.391304	-0.041703
Sud Ouest	23	0.521739	0.583840
Nord Ouest	21	0.238095	-0.741941
Sud Est	25	0.440000	0.200000
Total	100	0.400000	

Maximum Deviation Occurred at Observation 52
Value of sevesoHR at Maximum = 3.0

Kolmogorov-Smirnov Statistics (Asymptotic)
KS 0.096855 KSa 0.968548

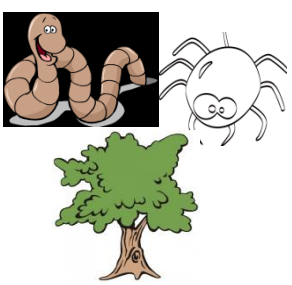


Statistiques non paramétriques : proc npar1way

Test	Probabilité
ANOVA	0.81
Médiane	0.70
Chi2	0.75
Van der Waerden	0.75
Savage	0.85
Kruskall–Wallis	0.77
Kolmogorov–Smirnov	0.97



Tout le monde est d'accord,
l'effet est non significatif

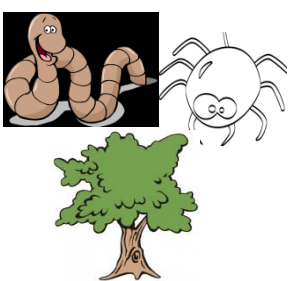


Statistiques non paramétriques : proc npar1way

Lire le fichier « Longevite.csv » qui contient :

- Le nom de l'espèce (26 caractères au maximum)
- Sa longevite moyenne en années
- Son règne
- Sa Catégorie
- Si on a un record de longévité

```
data longevite;  
: infile "FormationSAS/Longevite.csv" dlm=';' firstobs=2;  
: input espece :$26. longevite regne $ categorie :$10. record $;  
: ;  
: proc npar1way ; class categorie; var longevite;run;  
: ;
```



Exercice 5

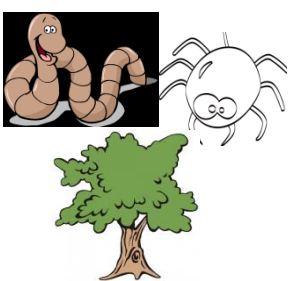
The NPAR1WAY Procedure

Analysis of Variance for Variable longevite
Classified by Variable categorie

categorie	N	Mean
Arthropode	33	12.64212
Oiseau	35	27.37143
Reptile	28	53.60714
Arbres	46	1952.41304
Mammifère	68	28.69118
Spongiaire	8	1905.37500
Poisson	26	24.03846
Mollusque	13	77.11538
Amphibien	12	15.91667

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Among	8	158272874.813	19784109.35	2.5863	0.0098
Within	260	1988920676.272	7649694.91		

Average scores were used for ties.



Exercice 5

The NPAR1WAY Procedure

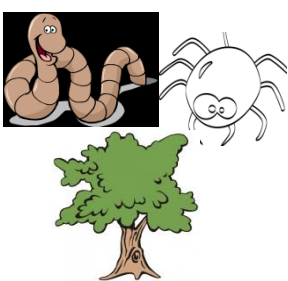
Wilcoxon Scores (Rank Sums) for Variable longevite
Classified by Variable categorie

categorie	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
Arthropode	33	1950.00	4455.0	418.369337	59.090909
Oiseau	35	4680.00	4725.0	429.031150	133.714286
Reptile	28	4152.00	3780.0	389.434490	148.285714
Arbres	46	10701.00	6210.0	480.151425	232.630435
Mammifère	68	8774.00	9180.0	554.241907	129.029412
Spongiaire	8	1247.50	1080.0	216.626769	155.937500
Poisson	26	2485.00	3510.0	376.822378	95.576923
Mollusque	13	1274.00	1755.0	273.488163	98.000000
Amphibien	12	1051.50	1620.0	263.271628	87.625000

Average scores were used for ties.

Kruskal-Wallis Test

Chi-Square 119.8672
DF 8
Pr > Chi-Square <.0001



Exercice 5

The NPAR1WAY Procedure

Median Scores (Number of Points Above Median) for Variable longevite
Classified by Variable categorie

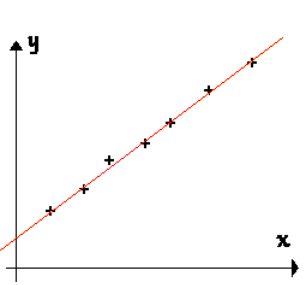
categorie	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
Arthropode	33	5.0	16.438662	2.695334	0.151515
Oiseau	35	15.0	17.434944	2.764022	0.428571
Reptile	28	17.0	13.947955	2.508922	0.607143
Arbres	46	45.0	22.914498	3.093363	0.978261
Mammifère	68	33.0	33.873606	3.570690	0.485294
Spongiaire	8	5.0	3.985130	1.395613	0.625000
Poisson	26	7.0	12.951673	2.427669	0.269231
Mollusque	13	3.0	6.475836	1.761941	0.230769
Amphibien	12	4.0	5.977695	1.696121	0.333333

Average scores were used for ties.

Median One-Way Analysis

Chi-Square 71.0493
DF 8
Pr > Chi-Square <.0001

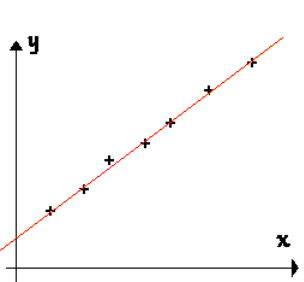




La régression linéaire : proc reg

- ▶ Régression linéaire simple. Toutes les variables du modèle sont testées simultanément

```
00022
00023 data seveso;
00024 infile "FormationSAS/DonneesSEVESO.csv" firstobs=2 dlm=';';
00025 input departement :$23. autorisation sevesoHR sevesoRM region :$13.
00026     pop2009 superficie taille :$10. densite densiteb :$19.;
00027 proc reg;
00028 model autorisation = sevesoHR sevesoRM pop2009 superficie densite ;
00029 run;
00030
00031
00032
```



La régression linéaire : proc reg

The REG Procedure
 Model: MODEL1
 Dependent Variable: autorisation

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	3709430	741886	4.05	0.0023
Error	94	17229929	183297		
Corrected Total	99	20939359			

Root MSE	428.13213	R-Square	0.1772
Dependent Mean	510.53000	Adj R-Sq	0.1334
Coeff Var	83.86033		

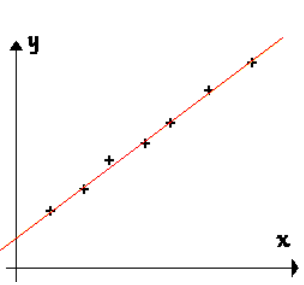
Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	231.25276	86.90605	2.66	0.0092
sevesoHR	1	-9.76738	9.40730	-1.04	0.3018
sevesoRM	1	1.19663	12.04077	0.10	0.9210
pop2009	1	0.00054940	0.00015318	3.59	0.0005
superficie	1	0.00151	0.00558	0.27	0.7878
densite	1	-0.05109	0.02444	-2.09	0.0393

Coefficients significativement différents de 0
 Seulement pour Intercept, pop2009 et densite

Estimation des coefficients de la régression linéaire et de leur erreur d'estimation :

Autorisation = 231.25 - 9.77 sevesoHR + 1.20 sevesoRM + 0.0005 pop2009 + 0.0015 superficie - 0.05 densite

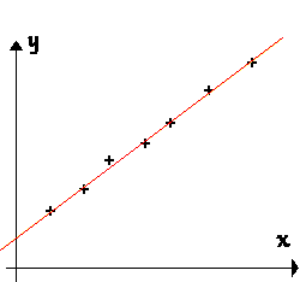


La régression linéaire : proc reg

- ▶ Régression linéaire par étape. Les variables du modèle sont introduites une par une, par ordre d'importance, jusqu'à ce que le R^2 de la régression n'augmente plus significativement

```
1
2
3 data seveso;
4 infile "FormationSAS/DonneesSEVESO.csv" firstobs=2 dlm=';';
5 input departement :$23. autorisation sevesoHR sevesoRM region :$13.
6     pop2009 superficie taille :$10. densite densiteb :$19.;
7 proc reg;
8 model autorisation = sevesoHR sevesoRM pop2009 superficie densite / selection=stepwise;
9 run;
0
1
```

Option indiquant l'introduction progressive des variables



La régression linéaire : proc reg



The SAS System

The REG Procedure

Model: MODEL1

Dependent Variable: autorisation

Stepwise Selection: Step 1

Etape 1 : introduction de la variable qui augmente le plus le R^2 (pop2009)

Variable pop2009 Entered: R-Square = 0.1358 and C(p) = 2.7255

Analysis of Variance

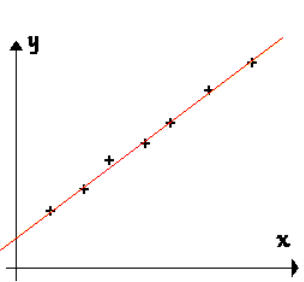
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	2843267	2843267	15.40	0.0002
Error	98	18096092	184654		
Corrected Total	99	20939359			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	287.39508	71.27457	3002259	16.26	0.0001
pop2009	0.00034772	0.00008861	2843267	15.40	0.0002

Bounds on condition number: 1, 1

Stepwise Selection: Step 2

Résultats de l'étape 1



La régression linéaire : proc reg

Etape 2 : introduction de la variable qui augmente le plus le R^2 après pop2009

Stepwise Selection: Step 2

Variable densite Entered: R-Square = 0.1664 and C(p) = 1.2235

Analysis of Variance

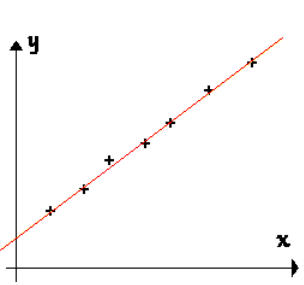
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	3485160	1742580	9.68	0.0001
Error	97	17454199	179940		
Corrected Total	99	20939359			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	251.33267	72.90368	2138588	11.88	0.0008
pop2009	0.00043599	0.00009918	3477427	19.33	<.0001
densite	-0.03821	0.02023	641892	3.57	0.0619

Bounds on condition number: 1.2854, 5.1418

All variables left in the model are significant at the 0.1500 level.

Résultats de l'étape 1



La régression linéaire : proc reg

Etape 3 : aucune variable ne permet d'augmenter significativement le R^2

Stepwise Selection: Step 2

No other variable met the 0.1500 significance level for entry into the model.

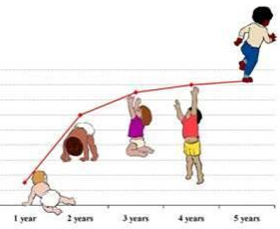
Summary of Stepwise Selection

Step	Variable Entered	Variable Removed	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	pop2009		1	0.1358	0.1358	2.7255	15.40	0.0002
2	densite		2	0.0307	0.1664	1.2235	3.57	0.0619

Le R^2 partiel donne l'augmentation de R^2 du modèle due à l'introduction de la variable
 Le R^2 du modèle donne le R^2 total

Résultats final

Régression non linéaire : proc nlin



```
data a;
infile "Cameroun-tarseemplume.csv" firstobs=3 dlm=';';
input p0m p0f p1m p1f p2m p2f p3m p3f p4m p4f p5m p5f p6m p6f
      p7m p7f p8m p8f p9m p9f p10m p10f p11m p11f p12m p12f
      p13m p13f p14m p14f p15m p15f p16m p16f ;
truc=1;
proc sort; by truc;
proc means noprint;
output out=meansa VAR=
      m0m m0f m1m m1f m2m m2f m3m m3f m4m m4f m5m m5f m6m m6f
      m7m m7f m8m m8f m9m m9f m10m m10f m11m m11f m12m m12f
      m13m m13f m14m m14f m15m m15f m16m m16f ;

data meansa2; set meansa; truc=1;proc sort; by truc;

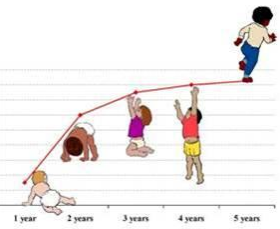
data total; merge a meansa2; by truc;

data j0m ;set total; age=0 ; sexe=1; poids=p0m ; pond=(m16m/m0m);
data j1m ;set total; age=7 ; sexe=1; poids=p1m ; pond=(m16m/m1m);
data j2m ;set total; age=14 ; sexe=1; poids=p2m ; pond=(m16m/m2m);
data j3m ;set total; age=21 ; sexe=1; poids=p3m ; pond=(m16m/m3m);
data j4m ;set total; age=28 ; sexe=1; poids=p4m ; pond=(m16m/m4m);
data j5m ;set total; age=35 ; sexe=1; poids=p5m ; pond=(m16m/m5m);
data j6m ;set total; age=42 ; sexe=1; poids=p6m ; pond=(m16m/m6m);
data j7m ;set total; age=49 ; sexe=1; poids=p7m ; pond=(m16m/m7m);
data j8m ;set total; age=56 ; sexe=1; poids=p8m ; pond=(m16m/m8m);
data j9m ;set total; age=63 ; sexe=1; poids=p9m ; pond=(m16m/m9m);
data j10m;set total; age=70 ; sexe=1; poids=p10m; pond=(m16m/m10m);
data j11m;set total; age=77 ; sexe=1; poids=p11m; pond=(m16m/m11m);
data j12m;set total; age=84 ; sexe=1; poids=p12m; pond=(m16m/m12m);
data j13m;set total; age=91 ; sexe=1; poids=p13m; pond=(m16m/m13m);
data j14m;set total; age=98 ; sexe=1; poids=p14m; pond=(m16m/m14m);
data j15m;set total; age=105; sexe=1; poids=p15m; pond=(m16m/m15m);
data j16m;set total; age=112; sexe=1; poids=p16m; pond=(m16m/m16m);
```

Le fichier doit contenir une ligne par individu et par temps.

La variable « pond » contient un facteur de pondération à utiliser dans la régression non linéaire afin de donner plus de poids aux variables de valeur numérique faible

Régression non linéaire : proc nlin



```
data j0f ;set total; age=0 ; sexe=2; poids=p0f ; pond=(m16f/m0f);
data j1f ;set total; age=7 ; sexe=2; poids=p1f ; pond=(m16f/m1f);
.....
data j15f;set total; age=105; sexe=2; poids=p15f; pond=(m16f/m15f);
data j16f;set total; age=112; sexe=2; poids=p16f; pond=(m16f/m16f);
```

```
data total2; set
j0m j1m j2m j3m j4m j5m j6m j7m j8m j9m j10m
j11m j12m j13m j14m j15m j16m
j0f j1f j2f j3f j4f j5f j6f j7f j8f j9f j10f
j11f j12f j13f j14f j15f j16f;
```

```
proc sort; by sexe ;
proc nlin; by sexe ;
parms p0c=20 to 60 by 10 B=2 to 8 by 2 K=0.01 to 0.10 by 0.01;
bounds B K>0, p0c<100;
model poids=p0c*exp(B*(1-exp(-K*age)));
_weight_=pond;
output out=sortie parms=p0c B K;

proc sort data=sortie; by sexe;

data sortie2;
set sortie; by sexe;
if first.sexe;
L=B*K;
TI=(1/K)*log(B);
A=p0c*exp(B);
proc print; var sexe L TI A;

run;
```

Le fichier total2 contient bien 1 ligne par âge et par individu

Points de départ de la modélisation

Bornes que les paramètres ne doivent pas dépasser

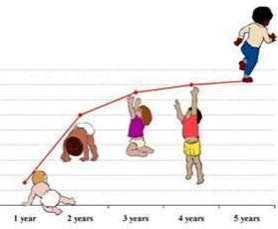
Modèle mathématique

Indication de la variable de pondération des observations

Récupération des résultats dans le fichier « sortie »

Pour éviter de trop longs listings, on peut utiliser l'option « noprint » de proc nlin : « *proc nlin noprint;* » au lieu de « *proc nlin;* ».

Régression non linéaire : proc nlin



```
SAS: Log-Untitled
File Edit View Tools Solutions Help

NOTE: PROCEDURE SORT used:
      real time          0.04 seconds
      cpu time           0.04 seconds

238 proc nlin; by sexe ;
239 parms p0c=20 to 60 by 10 B=2 to 8 by 2 K=0.01 to 0.10 by 0.01;
240 bounds B K>0, p0c<100;
241 model poids=p0c*exp(B*(1-exp(-K*age)));
242 _weight_=pond;
243 output out=sortie parms=p0c B K;

NOTE: DER.p0c not initialized or missing. It will be computed automatically.
NOTE: DER.B not initialized or missing. It will be computed automatically.
NOTE: DER.K not initialized or missing. It will be computed automatically.
NOTE: PROC NLIN grid search time was 0: 0: 1.
NOTE: Convergence criterion met.
NOTE: The above message was for the following by-group:
      sexe=1
NOTE: PROC NLIN grid search time was 0: 0: 1.
NOTE: Convergence criterion met.
NOTE: The above message was for the following by-group:
      sexe=2
NOTE: There were 2448 observations read from the data set WORK.TOTAL2.
NOTE: The data set WORK.SORTIE has 2448 observations and 78 variables.
NOTE: PROCEDURE NLIN used:
      real time          3.59 seconds
      cpu time           1.88 seconds
```

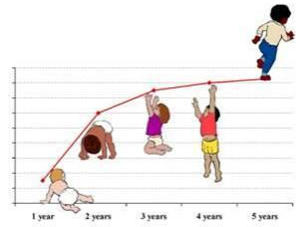
Penser à regarder dans la fenêtre de log si l'exécution du programme a bien fonctionné.

La procédure fournira quoiqu'il arrive des résultats, même si la procédure n'a pas réussi à trouver l'ajustement adéquat. Dans le log, on aura alors des messages d'erreur du genre « convergence failed »

```
NOTE: PROC NLIN grid search time was 0: 0: 1.
NOTE: Convergence criterion met.
NOTE: The above message was for the following by-group:
      sexe=2
NOTE: PROC NLIN grid search time was 0: 0: 1.
WARNING: PROC NLIN failed to converge.
NOTE: The above message was for the following by-group:
      sexe=3
NOTE: There were 2454 observations read from the data set W
```

Exemple en introduisant des données fictives et aberrantes pour un 3^{ème} sexe

Régression non linéaire : proc nlin



Courbe de Gompertz

3702
09:24 Thursday, March 31, 2011

----- sexe=2 -----

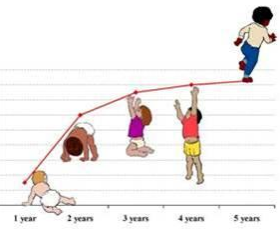
The NLIN Procedure
Grid Search
Dependent Variable poids

p0c	B	K	Weighted SS
20.0000	2.0000	0.0100	2.4706E8
30.0000	2.0000	0.0100	2.0572E8
40.0000	2.0000	0.0100	1.7852E8
50.0000	2.0000	0.0100	1.6547E8
60.0000	2.0000	0.0100	1.6656E8
20.0000	4.0000	0.0100	1.3615E8
30.0000	4.0000	0.0100	66415152
40.0000	4.0000	0.0100	28875805
50.0000	4.0000	0.0100	23527039

Somme des carrés à minimiser

Points de départ de la procédure

Régression non linéaire : proc nlin



----- sexe=2 -----

The NLIN Procedure
Iterative Phase
Dependent Variable poids
Method: Gauss-Newton

Iter	p0c	B	K	Weighted SS
0	20.0000	4.0000	0.0200	14310566
1	32.2877	3.6164	0.0161	9278582
2	32.3934	3.7576	0.0168	3582675
3	32.4076	3.7441	0.0168	3537227
4	32.4070	3.7440	0.0168	3537225

Le meilleur point de départ est p0c à 20, B à 4 et K à 0.02

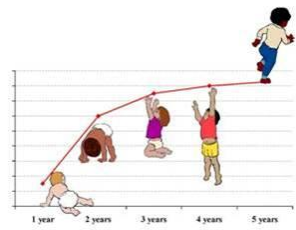
Ensuite, on affine

NOTE: Convergence criterion met.

Le résultat est positif !

Estimation Summary

Method	Gauss-Newton
Iterations	4
R	4.295E-6
PPC(K)	1.955E-6
RPC(K)	0.00004
Object	5.279E-7
Objective	3537225



Régression non linéaire : proc nlin

----- sexe=2 -----

The NLIN Procedure

Source	DF	Sum of Squares	Mean Square	F Value	Approx Pr > F
Regression	3	3.6864E8	1.2288E8	37900.3	<.0001
Residual	1091	3537225	3242.2		
Uncorrected Total	1094	3.7218E8			
Corrected Total	1093	2.5623E8			

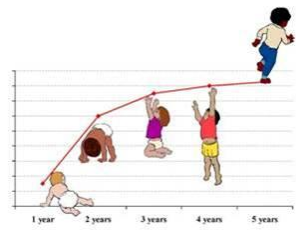
Le modèle d'ajustement est-il pertinent ?

Parameter	Estimate	Approx Std Error	Approximate 95% Confidence Limits	
p0c	32.4070	0.2351	31.9458	32.8683
B	3.7440	0.0215	3.7019	3.7861
K	0.0168	0.000231	0.0163	0.0172

Estimation des paramètres et de leur intervalle de confiance

Approximate Correlation Matrix			
	p0c	B	K
p0c	1.0000000	-0.0656612	-0.3677694
B	-0.0656612	1.0000000	-0.8838411
K	-0.3677694	-0.8838411	1.0000000

Estimation des corrélations entre paramètres



Régression non linéaire : proc nlin

```

data fin;
set resultats; by poulet;
if first.poulet;
L = B* K;
TI=(1 /K)*log(B);

proc glm; class sexe;
model P0e L K TI = sexe;
lsmeans sexe;

run;

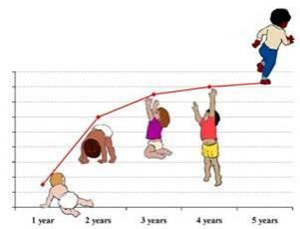
```

Pour éviter de trop longs listings, on peut utiliser l'option « noprint » de proc nlin :
 « `proc nlin noprint;` » au lieu de « `proc nlin;` ».

Attention cependant à regarder dans la fenêtre de log si l'exécution du programme a bien fonctionné.

La procédure fournira quoiqu'il arrive des résultats, même si la procédure n'a pas réussi à trouver l'ajustement adéquat. Dans le log, on aura alors des messages d'erreur du genre « convergence failed »

Pour récupérer les résultats, il faut recréer un fichier supplémentaire « fin », dans lequel on met le fichier « resultats », en récupérant seulement la 1^{ère} ligne de chaque animal (by poulet). On peut ensuite recréer de nouvelles variables (L : vitesse de croissance initiale ou TI : âge au point d'inflexion), puis les analyser.



Régression non linéaire : proc nlin

Iter	p0c	B	K	Weighted SS
96	100.0	51.9433	0.000249	172771
97	100.0	52.4231	0.000247	172745
98	100.0	53.0314	0.000244	172724
99	100.0	53.8055	0.000240	172714
100	100.0	54.3003	0.000238	172690

WARNING: Maximum number of iterations exceeded.



WARNING: PROC NLIN failed to converge.



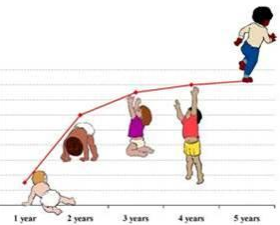
----- sexe=3 -----

The NLIN Procedure

Source	DF	Sum of Squares	Mean Square	F Value	Approx Pr > F
Regression	2	1064110	532055	2.86	0.1663
Residual	4	172690	43172.6		
Uncorrected Total	6	1236800			
Corrected Total	5	296000			



Régression non linéaire : proc nlin



Parameter	Estimate	Approx Std Error	Approximate 95% Confidence Limits		Label
p0c	100.0	0	100.0	100.0	
B	54.3003	5508.5	-15239.6	15348.2	
K	0.000238	0.0245	-0.0678	0.0682	
Bound2	1973.0	460.0	695.8	3250.1	p0c < 100

Courbe de Gompertz

09:24 Thursday, March 31, 2011 3771

Obs	sexe	L	TI	A
1	1	0.069600	84.69	2071.59
2	2	0.062764	78.75	1369.74
3	3	0.012931	16773.84	3.8222E25



Exercice 4

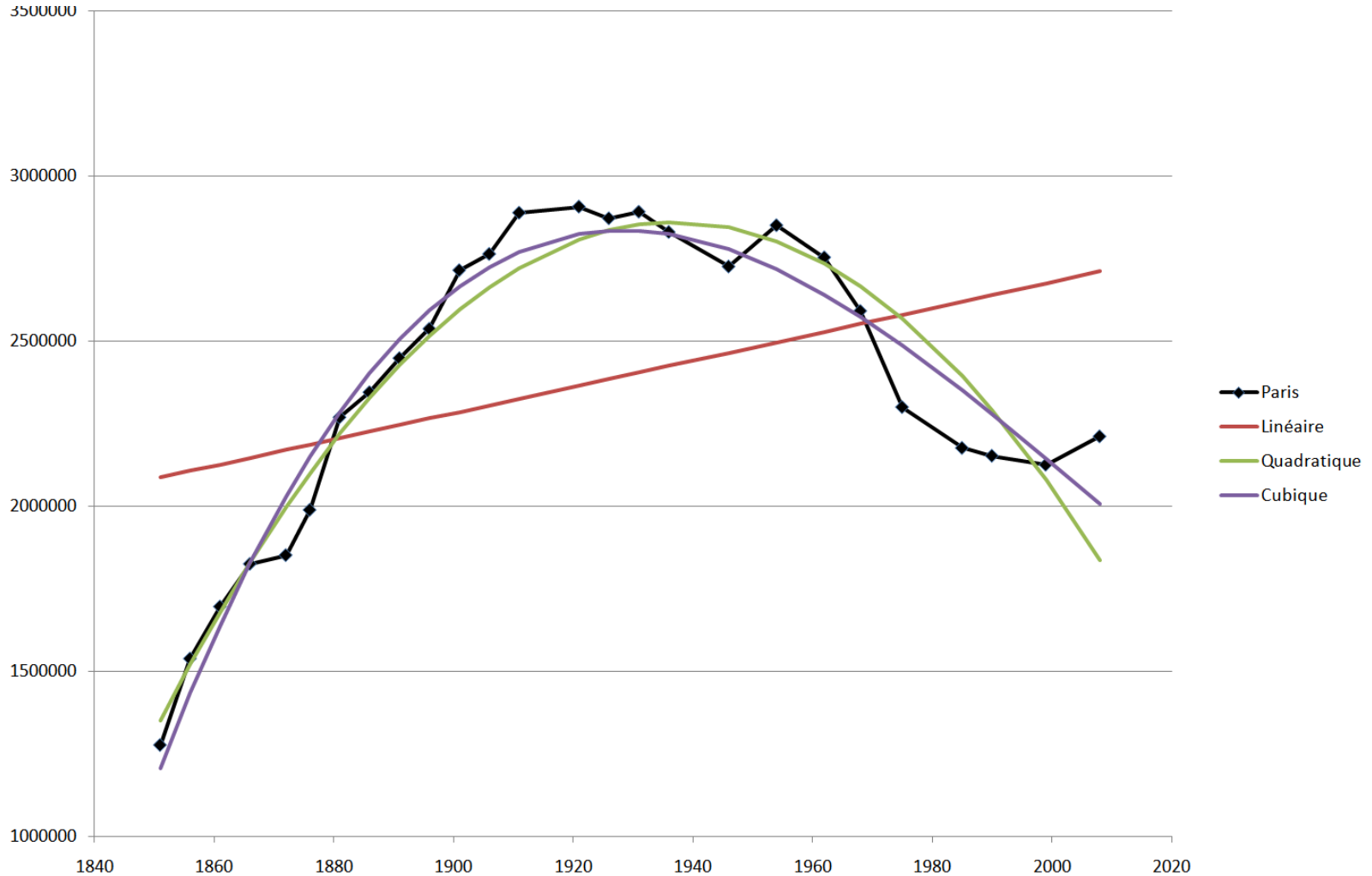
- ▶ Reprendre le fichier de Données de populations et faire
 - Une régression linéaire de la population de Paris sur l'année
 - Une régression quadratique, cubique

```

} proc reg; model DEP75 = annee;
}
} proc nlin;
}   parms a=1 to 100 by 50 b=0 to 100 by 50 c=0 to 100 by 50 ;
}   model Dep75=a+b*annee+c*(annee**2);
}
}
} proc nlin;
}   parms a=1 to 100 by 50 b=0 to 100 by 50 c=0 to 100 by 50 D=0 to 100 by 50;
}   model Dep75=a+b*annee+c*annee**2+d*annee**3;
}
}

```

Exercice4



PENGUINS ARE BLACK AND WHITE.
SOME OLD TV SHOWS ARE BLACK AND WHITE.
THEREFORE, SOME PENGUINS ARE OLD TV SHOWS.



Correlation is not causation

Corrélations entre variables : proc corr

SAS: Program Editor–Untitled–PROC REG running

File Edit View Tools Run Solutions Help

```
0031 data seveso;  
0032 infile "FormationSAS/DonneesSEVESO.csv" firstobs=2 dlm=';';  
0033 input departement :$23. autorisation sevesoHR sevesoRM region :$13.  
0034      pop2009 superficie taille :$10. densite densiteb :$19.;  
0035 proc corr;  
0036 var autorisation sevesoHR sevesoRM pop2009 superficie densite ;  
0037 proc corr spearman;  
0038 var autorisation sevesoHR sevesoRM pop2009 superficie densite ;  
0039 run;
```

Proc corr **sans option**



Corrélation paramétrique de Pearson

Proc corr **avec option spearman**



Corrélation non paramétrique de Spearman
(corrélation de rang)

PENGUINS ARE BLACK AND WHITE.
SOME OLD TV SHOWS ARE BLACK AND WHITE.
THEREFORE, SOME PENGUINS ARE OLD TV SHOWS.



Corrélations entre variables : proc corr

Correlation is not causation

Pearson Correlation Coefficients, N = 100
Prob > |r| under H0: Rho=0

	autorisation	sevesoHR	sevesoRM	pop2009	superficie	densite
autorisation	1.00000	0.21743 0.0298	0.23138 0.0205	0.36849 0.0002	-0.03929 0.6980	0.01922 0.8495
sevesoHR	0.21743 0.0298	1.00000	0.69807 <.0001	0.57985 <.0001	0.10690 0.2898	-0.09691 0.3375
sevesoRM	0.23138 0.0205	0.69807 <.0001	1.00000	0.51954 <.0001	0.05590 0.5807	-0.09147 0.3654
pop2009	0.36849 0.0002	0.57985 <.0001	0.51954 <.0001	1.00000	-0.15130 0.1329	0.47123 <.0001
superficie	-0.03929 0.6980	0.10690 0.2898	0.05590 0.5807	-0.15130 0.1329	1.00000	-0.14534 0.1491
densite	0.01922 0.8495	-0.09691 0.3375	-0.09147 0.3654	0.47123 <.0001	-0.14534 0.1491	1.00000

Valeur de la corrélation

**Probabilité associée à la corrélation
($<0.05 \Rightarrow$ significativement différente de 0)**

PENGUINS ARE BLACK AND WHITE.
SOME OLD TV SHOWS ARE BLACK AND WHITE.
THEREFORE, SOME PENGUINS ARE OLD TV SHOWS.



Correlation is not causation

Corrélations entre variables : proc corr

Spearman Correlation Coefficients, N = 100
Prob > |r| under H0: Rho=0

	autorisation	sevesoHR	sevesoRM	pop2009	superficie	densite
autorisation	1.00000	0.48002 <.0001	0.47193 <.0001	0.63446 <.0001	0.30395 0.0021	0.43716 <.0001
sevesoHR	0.48002 <.0001	1.00000	0.61044 <.0001	0.53743 <.0001	0.40148 <.0001	0.35682 0.0003
sevesoRM	0.47193 <.0001	0.61044 <.0001	1.00000	0.55891 <.0001	0.26556 0.0076	0.43682 <.0001
pop2009	0.63446 <.0001	0.53743 <.0001	0.55891 <.0001	1.00000	-0.05765 0.5689	0.89416 <.0001
superficie	0.30395 0.0021	0.40148 <.0001	0.26556 0.0076	-0.05765 0.5689	1.00000	-0.37686 0.0001
densite	0.43716 <.0001	0.35682 0.0003	0.43682 <.0001	0.89416 <.0001	-0.37686 0.0001	1.00000

Valeur de la corrélation

**Probabilité associée à la corrélation
($<0.05 \Rightarrow$ significativement différente de 0)**



Exercice 2



- Calculer les corrélations entre les différentes variables du fichier céréales
- Sauvegarder les résultats sous format excel par la procédure « PROC EXPORT »

```
proc corr data=cerealesmerge outs=correlation;
var calorie protides lipides glucides sucre taux_cereale sel;
```

```
proc export data=correlation outfile="FormationSAS/Exercices/CorrelationsCereales.csv";
run;
```

	<u>_NAME_</u>						
<u>_TYPE_</u>	calorie	protides	lipides	glucides	sucre	taux_cereale	sel
MEAN	389.150442	7.60265487	6.30176991	75.4571429	26.7052632	63.82564103	0.389726027
STD	31.9095321	1.55071759	5.7595415	9.26462529	10.194023	17.07364823	0.255375392
N	113	113	113	112	76	78	73
CORR	calorie	protides	lipides	glucides	sucre	taux_cereale	sel
calorie	1.00	-0.27	0.59	-0.10	0.51	-0.67	0.01
protides	-0.27	1.00	0.28	-0.65	-0.61	0.18	-0.24
lipides	0.59	0.28	1.00	-0.81	0.21	-0.65	-0.54
glucides	-0.10	-0.65	-0.81	1.00	0.19	0.30	0.57
sucre	0.51	-0.61	0.21	0.19	1.00	-0.67	-0.15
taux_cereale	-0.67	0.18	-0.65	0.30	-0.67	1.00	0.33
sel	0.01	-0.24	-0.54	0.57	-0.15	0.33	1.00
Corrélations positives significativement différentes de 0							
Corrélations négatives significativement différentes de 0							