

```

*/ Exercice 1;
data note;
input nom :$ int1 int2 int3;
lines;
Eric 12 3 8
Maxence 15 4 13
Margaux 12 13 17
Xavier 4 12 10
Elisa 16 5 9
;
run;
data note; set note;
etendue=max(of int1-int3)-min(of int1-int3);
moyenne=mean(of int1-int3);
std=std(of int1-int3);
format etendue 4.2 moyenne 4.2 std 4.2;
run;
proc print data=note;run;
proc sort data=note out=note;
by nom; run;
proc sort data=note out=note;
by descending moyenne; run;
proc sort data=note out=note;
by descending int2;
run;
proc print data=note;
var int2;
id nom;
run;

```

```

*/ Exercice 2;

```

```

*/ On commence par importer les données
temperatures.csv;
proc import datafile="C:\cours\sas\stage-sas
\donnees\temperatures.xls"
out=work.temp
dbms=excel replace;
getnames=yes;
run;

```

```

*/ Question 2;
data temp; set temp;
format year 4.0 month 2.0;

```

```

run;
*/ Question 3;
data temp; set temp;
/* dans la fonction MDY les trois arguments sont
nécessaires */
/* comme on n'a pas de variable Jour, on met une
valeur quelconque */
date=mdy(month,1,year);
/* dans l'affichage de la date, on utilise un
format qui reprend uniquement l'année et le
mois*/
format date monyy5.;
run;
*/ Question4 ;
proc print data=temp (obs=10);
var precp temp;
id date;
run;
*/ Question 5;
proc means data=temp noprint;
var precp;
output out=tab1 mean=m1 std=sigma max=maxi
min=mini;
run;
proc print data=tab1;run;
*/ Question 6 ;
proc means data=temp noprint;
var temp precp;
class year;
output out=tab2 mean=m_temp m_precp;
run;
data tab2; set tab2;
if _type_=0 then delete;
drop _type_;run;
proc print data=tab2;
var m_temp m_precp;
id year;run;
symbol1 i=join v=none c=black;
symbol2 i=join v=none c=red;
proc gplot data=tab2;
plot m_temp*year=1 m_precp*year=2 / overlay;
title 'Moyennes annuelles des températures et des
précipitations';
run;

```

```

/* Avec les deux échelles en ordonnée */
proc gplot data=tab2;
plot m_temp*year=1;
plot2 m_precp*year=2;
title 'Moyennes annuelles des températures et des
précipitations';
run;
*/ Question 7;
proc means data=temp noprint;
var temp precp;
class month;
output out=tab3 mean=m_temp m_precp;
run;
data tab3; set tab3;
if _type_=0 then delete;
drop _type_;run;
proc print data=tab3;
var m_temp m_precp;
id month;run;
proc gplot data=tab3;
plot m_temp*month=1 m_precp*month=2 / overlay;
title 'Moyennes mensuelles des températures et
des précipitations';
run;
/* Avec les deux échelles en ordonnée */
proc gplot data=tab3;
plot m_temp*month=1;
plot2 m_precp*month=2;
title 'Moyennes mensuelles des températures et
des précipitations';
run;

*/ Exercice 3;
symbol1 i=join v=none c=black;
symbol2 v=star c=red L=2;
proc gplot data=temp;
plot precp*date=1 temp*date=2 / overlay;
title 'Précipitations et températures au Canada';
run; quit;

/* Exercice 4*/
/* Importation des données */
/* La proc IMPORT est l'équivalent du menu

```

```

"Import data" */
PROC IMPORT OUT= WORK.CARS
            DATAFILE= "C:\cours\sas\stage-sas
\donnees\cars.xls"
            DBMS=EXCEL5 REPLACE;
            GETNAMES=YES;
RUN;

/* Formatage des variables numériques */
data cars; set cars;
format price 8.2 mileage 6.0 cylinder 1.0 liter
3.1 doors 1.0 cruise 1.0 sound 1.0 leather 1.0;
run;

/* Q2. Transformer la variable contenant le
nombre de miles en kilomètres */
data cars; set cars;
kilometrage=mileage*1.6;
drop mileage;run;

/* Q3. Représenter le nuage des points en
utilisant les variables de prix
et de kilometrage */
symbol1 i=none v=star c=red;
proc gplot data=cars;
plot price*kilometrage=1;
run;

/* A première vue, il n'y a pas de lien entre le
prix et le kilomètre!!! */

/* Q4. Isoler les voitures dont le prix est
supérieur à 52.000$. De quelle marque
et modèle s'agit-il? */
proc print data=cars; where price>52000;
run;

/* Il ne s'agit donc pas de valeurs
"aberrantes"!!!!
/* Plus tard, réaliser une régression linéaire
sur les 10 voitures en question*/
/* En cours d'analyse de données, faire des
classifications, de la sélection
de variables, tester plusieurs modèles de
régression, etc.... */

```

```

quit;

/* Exercice 5
/* Importation du fichier des données */
filename f 'C:\cours\sas\stage-sas\donnees
\nascar.txt';
data nascar; infile f lrecl=100 pad;
input Series 1-3 Year 6-9 @12 Race_Year 2.0 @16
Finish_Pos 2.0 @20 Start_Pos 2.0
@24 Laps 3.0 @29 Winnings 7.0 @38 Cars 2.0 @42
Car_Make $9. @53 Driver $30.;
run;
proc contents data=nascar;run;

/
*****/

/* Q2. Calculer la somme des gains par année et
par pilote */
proc means data=nascar noprint;
by year; class driver;
var winnings;
output out=gain_annuel sum=gain;
run;
data gain_annuel; set gain_annuel;
if _type_=0 then delete; drop _type_;
run;
/* Analyse de la distribution avec SAS insight */

/* Q3. Calculer le total des gains par année.
Faire une représentation graphique avec
SAS Insight. Quel type d'ajustement vous semble
pertinent? */
proc means data=gain_annuel noprint;
by year;
var gain;
output out=gain_total sum=gain;
run;

goptions reset=title;
symbol1 i=none v=star c=red;
proc gplot data=gain_total;

```

```

plot gain*year=1;
run;

/* Q4. Calculer la proportion des pilotes qui ont
fini la course */
/* On construit une nouvelle base qui ne contient
que les pilotes ayant fini la course */
data all_laps; set nascar;
retain test 0;
if Finish_Pos=1 then test=Laps;
if Laps<test then delete;
run;
proc contents data=nascar;run;
proc contents data=all_laps;run;

/* Exercice 6
/* Importation du fichier des données */
filename f 'C:\cours\sas\stage-sas\donnees
\electricbill.txt';
data electric; infile f;
input NUM 1-3 YEAR 5-8 MONTH $ 10-12 BILL 14-19
TEMP 21-24 HDD 26-29 CDD 31-33 SIZE 35 METER 37
PUMP1 39 PUMP2 41 RIDER_TOTAL 43-52 CONSUMPTION
54-58;
run;
proc contents data=electric;run;

/
*****
*****/

/* Q2. Transformer les degrés F en degrés °C */
data electric; set electric;
temp_C=(5/9)*(temp-32);
format temp_C 4.1;
drop temp;
run;

/* Q3. Créer une variable date contenant le mois
et l'année */
/* !!!! Il n'existe pas d'informat SAS pour le
mois uniquement */
data electric; set electric;
a=mod(num,12);

```

```
if a=0 then a=12;
ident=mdy(a,1,year);
format ident mmyys10.;
drop a;
run;
```

```
/* Q4. Représenter graphiquement la série des factures */
```

```
symbol1 i=join v=none c=black;
proc gplot data=electric;
plot bill*ident=1/skipmiss;
run;
```

```
/* Une erreur de saisie apparaît en décembre 1992*/
```

```
data electric; set electric;
if num=24 then year=1992;
a=mod(num,12);
if a=0 then a=12;
ident=mdy(a,1,year);
format ident mmyys10.;
drop a;
run;
```

```
/* On refait le graphique */
```

```
/* Changement du compteur lorsque la consommation devient nulle! */
```

```
/* Q5. Calculer et afficher les moyennes annuelles et mensuelles des factures */
```

```
proc means data=electric;
var bill; class year; run;
```

```
proc means data=electric;
var bill; class month; run;
```

```
/* Q6. Calculer et afficher les moyennes en fonction du nombre de personnes dans le foyer */
```

```
proc means data=electric;
var bill; class size; run;
```

```
/* Q7. Calculer et afficher les factures moyennes avant et après l'installation du nouveau compteur */
```

```

proc means data=electric;
var bill; class meter; run;

/* Exercice 7
/* Importation du fichier des données */
filename f 'C:\cours\sas\stage-sas\donnees
\tips.txt';
data tips; infile f;
input IDEN $ 1-4 TOTBILL 7-11 TIP 15-18 SEX 23-23
SMOKER 30-30 DAY 35-35 TIME 40-40
SIZE 45-45;
run;

/
*****
*****/

/* Q2. Créer une nouvelle variable qui représente
le pourcentage du pourboire dans l'addition */
data tips; set tips;
rate=(tip*100)/totbill;
format rate 5.2;
run;

/* Q3. Calculer des statistiques descriptives
pour les variables quantitatives */
proc means data=tips; var totbill tip rate size;
run;

/* Q4. Formater les variables day et time */
proc format;
value jour 3='Jeudi' 4='Vendredi' 5='Samedi'
6='Dimanche';
value moment 0='Dejeuner' 1='Diner';
run;
data tips; set tips;
format day jour. time moment.;
run;

/* Q5. Calculer le taux moyen pour le pourboire
par jour et par moment de la journée */
proc sort data=tips; by day;run;
proc means data=tips;
var tip;

```



```
by day;  
class time;  
run;
```

```
proc means data=tips;  
var rate;  
by day;  
class time;  
run;
```

**/* Q6. Même question selon le sexe, les fumeurs
et le nombre de personnes à table */**

```
proc sort data=tips; by sex;run;  
proc means data=tips;  
var rate;  
by sex;  
class smoker;  
run;  
proc sort data=tips; by size;run;  
proc means data=tips;  
var rate;  
class size;  
run;
```

**/* Q7. Tracer l'histogramme de la variable
pourboire avec SAS Insight et
en utilisant des amplitudes différentes.
Commenter les résultats. */**

```
proc univariate data=tips;  
var tip;  
histogram;  
run;
```

**/* Q8. Tracer l'histogramme de la variable
pourcentage du pourboire avec SAS Insight.
Identifier les valeurs extrêmes. A quel type de
consommateurs correspondent-elles?
Commenter les résultats. */**

```
proc univariate data=tips;  
var rate;  
histogram;  
run;
```

```

/* Q9. Représenter graphiquement le nuage de
points en croisant l'addition et le pourboire.
Que remarquez-vous */
symbol1 i=none v=star c=red;
proc gplot data=tips;
plot tip*totbill=1;
run;
/* Q10. Même question en séparant par sexe et
ensuite par fumeur/nonfumeur */

/* Exercice 8
/* Importation du fichier des données */
filename f 'C:\cours\sas\stage-sas\donnees
\poverty.txt';
data poverty; infile f lrecl=100 pad;
input Birth 1-6 Death 7-14 Infant 15-22 LE_Males
23-30 LE_Females 31-38
GNP 39-46 Group 47-52 Country $ 53-74;
run;
proc contents data=poverty;run;

/
*****/

/* Q1. Supprimer la dernière ligne */
data poverty; set poverty;
if _N_=98 then delete;
run;

/* Q2. Est-ce que la richesse du monde est
distribuée de manière équitable? */
proc sort data=poverty;
by group;run;
proc boxplot data=poverty;
plot gnp*group;
run;
/* Avec les outliers */
proc boxplot data=poverty;
plot gnp*group / boxstyle=schematic;
run;
/* Identifier les valeurs aberrantes; expliquer
les résultats */

```

```

/* Faire quelques exemples avec SAS Insight
(boxplots, histogrammes, tests
sur les moyennes, etc... )*/

/* A faire avec SAS INSIGHT uniquement */

/* Q3. Est-ce que l'espérance de vie est
similaire dans les différents pays? */

/* Q4. Est-ce que les hommes et les femmes ont
des espérances de vie similaires? */
/* Commenter et faire un test! */
/* Calculer la différence; quelle est la
différence moyenne? Minimum? Maximum?
Dans quels pays cela se produit? */
data poverty; set poverty;
le_diff=le_males-le_females;
run;
proc univariate data=poverty;
var le_diff; run;

/* Q5. Est-ce la natalité est liée à la
mortalité? */
/* Tracer le nuage de points avec Insight */

/* Q6. Quel est le taux de croissance de la
population? */
data poverty; set poverty;
taux=birth-death; run;
/* Avec SAS Insight, calculer le taux moyen, le
minimum, le maximum */
/* Identifier les pays extrêmes */
/* A ce rythme, en combien d'années, les pays
extrêmes verront leurs
populations diminuer de la moitié ou doubler? */
/* Répéter l'étude par groupe de pays */

/* Exercice 9
/* Importation du fichier des données */
filename f 'C:\cours\sas\stage-sas\donnees
\tdf.txt';
options yearcutoff=1900;
data tdf; infile f;
input year 1-5 start_town $ 6-24 @25 start_date

```

```

mmddy8. @34 end_date mmddy8.
Stages 43-44 Distance 46-49 Speed 51-55 Time
57-62 Winner $ 64-83
Country $ 84-95 @96 Birth_date mmddy8. age_year
105-106 age_tenth 108-111;
format start_date mmddy8. end_date mmddy8.
Birth_date mmddy8.;
run;
proc contents data=tdf;run;

/
*****
*****/

/* Q2. Vérifier si les variables vitesse,
distance et temps sont correctement renseignées .
On supposera que, par défaut, les variables mieux
renseignées sont le
temps et la distance */
data tdf; set tdf;
vitesse=distance/time;
diff=abs(vitesse-speed);
format vitesse 5.2 diff 5.2;
run;
proc sort data=tdf;
by descending diff;
run;
/* Une seule valeur doit être modifiée */
data tdf; set tdf;
if diff>0.5 then speed=distance/time;
format speed 5.2;
drop diff vitesse;run;

/* Q3. Vérifier si l'âge du gagnant est
correctement renseignée */
data tdf; set tdf;
year_birth=year(birth_date);
age=year-year_birth;
diff=abs(age-age_year);
run;
proc sort data=tdf;
by descending diff;
run;
/* !!!!! Problème avec l'année 2000 !!!!!*/

```

```

/* Il faut modifier les options par défaut */
data tdf; set tdf;
year_birth=year(birth_date);
age=year-year_birth;
diff=abs(age-age_year);
run;
proc sort data=tdf;
by descending diff;
run;
data tdf; set tdf;
drop year_birth age diff;run;

/* Q4. Représenter graphiquement la série des
âges des gagnants */
/* Etudier la distribution de cette variable */
/* On peut se servir de SAS Insight */
proc sort data=tdf;
by year;run;
symbol1 i=join v=star color=blue;
proc gplot data=tdf;
plot age_year*year=1;run;

/* Q5. Mêmes questions pour la distance totale et
la vitesse moyenne.
Que remarquez-vous? */
symbol1 i=join v=star color=blue;
proc gplot data=tdf;
plot distance*year=1;run;

symbol1 i=join v=star color=blue;
proc gplot data=tdf;
plot speed*year=1;run;

```