

```
/* Practica SAS */  
libname prac "C:\Users\aulas\Desktop";
```

915

```
Proc template;  
Define style Style.Custom;  
parent=styles.default;  
style usertext from usertext /  
    font_size=18pt  
    font_weight=bold;  
end;  
run;
```

```
ods pdf file='C:\Users\aulas\Desktop\Practica.pdf'  
    style=Style.Custom  
    title = 'Practica de SAS'  
    author = 'C...'  
    subject = 'Computación'  
    ;
```

ok (1)

```
/* Ej1. Lectura dades */  
PROC IMPORT OUT= prac.dades  
    DATAFILE= "C:\Users\aulas\Desktop\dades_finals_v2.xls"  
    DBMS=XLS REPLACE;  
    GETNAMES=YES;  
RUN;  
proc contents data = prac.dades; run;
```

```
/* Ej2.Etiquetas */  
proc format library = prac;  
value prov 8='Barcelona' 43='Tarragona' 17='Girona' 25='Lleida';  
value sino 2= 'No' 1 = 'Si';  
value lloc 1= 'Centre sanitari' 2 = 'Domicili particular' 3 = 'Un altre lloc';  
value norm 1="Normal" 2= "Complicacions";  
value prem 1="A termini" 2="Prematur";  
value estudis 0=" " 1,2,3="Sense estudis" 4,5="ESO/Batxillerat" 6,7="Cicle formatiu"  
    8,9,10="Universitat";  
value grand 1,2,3,4="Menys 100.000" 5= "Més de 100.000" 6="Capital de provincia";  
value sexe 1="Home" 6="Dona";  
run;
```

ok (1)

```
/* En el siguiente data/set se utiliza para realizar: 1) el rename de algunas variables,  
2) pasar a categóricas las variables  
correspondientes (put), 3) aplicar los formatos definidos con anterioridad*/  
options fmtsearch = (prac);  
data prac.dades;  
set prac.dades (rename = (Anys_de_relacio_estable = Anys_relacio Edat_de_la_mare=Edat_mare  
    Edat_del_pare=Edat_pare Num_de_nescuts_amb_o_sense_vida= Numero_nascuts  
    Num_de_setmanes_de_embaras=Setmanes_embaras Pes_del_Nado=Pes));  
  
Provincia = put(Provincia_del_part,prov.);  
Prematur = put(A_termini_o_prematur,prem.);  
Grandaria = put(Grandaria_municipi_naixement_de,grand.);  
Lloc_part = put(Lloc_del_part,lloc.);  
Nacionalitat = put(Nacionalitat_espanyola_de_la_mar,sino.);  
Estudis = put(Nivell_d_estudis_de_la_mare,estudis.);  
Cesaria = put(Part_amb_cesaria,sino.);  
Complicacions = put(Part_normal_o_sense_complicacion,sino.);  
Sexe = put(Sexe_del_nascut ,sexe.);
```

```
drop Provincia_del_part A_termini_o_prematur Grandaria_municipi_naixement_de Lloc_del_part
Nacionalitat_espanyola_de_la_mar Nivell_d_estudis_de_la_mare Part_amb_cesaria
Part_normal_o_sense_complicacion Sexe_del_nascut;
run;
```

```
/* 3.Descriptivo outliers*/
```

```
/* Se utiliza el proc univariate para estudiar los outliers y graficar histogramas*/
```

```
title 'Histograma de las variables numéricas';
```

```
proc univariate data=prac.dades noprint ;
```

```
histogram;
```

```
run;
```

```
title;
```

```
title 'Estadísticos para detercar outliers';
```

```
proc univariate data=prac.dades ;
```

```
run;
```

```
title;
```

```
/* Existen observaciones extremas especialmente en el peso del nacido y la edad del padre.*/
```

```
/* Ej.4. Justificació dels outliers*/
```

```
/* Los posibles valores extremos encontrados en el apartado anterior son casos que pueden ser factibles. No parece que sean
```

```
errores en la entrada de datos sino más bien situaciones excepcionales que escapan de la normalidad de la situación.
```

```
Consecuentemente se ha decidido no hacer ningún tipo de cambio en los datos.
```

```
Al tratarse de datos provenientes de organismos
```

```
oficiales era esperable que tuvieran una cierta calidad*/
```

```
/* Ej. 5 Descriptiva completa */
```

```
/* 5.1 Descriptiva para la variable edad de la madre*/
```

```
proc univariate data=prac.dades ;
```

```
var Edat_mare;
```

```
histogram;
```

```
run;
```

```
/* 5.2 Descriptiva para el peso y las semanas de embarazo*/
```

```
PROC MEANS DATA = prac.dades printalltypes mean median var min max skew kurt p5 q3 maxdec = 2;
```

```
var Pes Setmanes_embaras;
```

```
RUN;
```

```
/* 5.3 'Descriptiva para el nivel de estudios de la madre*/
```

```
PROC FREQ DATA = prac.dades;
```

```
table Estudis ;
```

```
RUN;
```

```
/* Entre las variables estudiadas, se observa como la edad media de las mujeres al embarazo es de 32.2 años, se está
```

```
de media 39 semanas de embarazo y el peso de la criatura es de 3222 gramos.
```

```
En cuanto a los estudios se observa como
```

```
el 40% de las madres tienen estudios universitarios o superiores mientras
```

```
que el 13% no tienen estudios.*/
```

```
/* Ej. 6 Agrupaments numeriques */
```

```
data prac.dades;
```

```
set prac.dades;
```

```
if Edat_mare<20 then Edat_Rang = "<20";
```

```
else if 20<= Edat_mare <= 27 then Edat_rang = "20-27";
```

```
else if 28<= Edat_mare <= 35 then edat_rang = "28-35";
```

ok (1)

ok (1)

ok (1)

ok (0.75)

Pero mejor hacerlo numérico formateado.

```

else if 35 <= edat_mare <= 42 then edat_rang = "35-42";
else edat_rang = ">42";
run;

```

```

PROC FREQ DATA = prac.dades;
table Edat_Rang ;
RUN;

```

```

/* Ej. 7*/
/* Utilizamos todas las variables con formato numérico en prac.dades y
calculamos las correlaciones 2 a 2 de Pearson y de Spearman.*/

```

```

proc iml;
use prac.dades;
read all var _NUM_ into X[colname=varNames];
corr = corr(X);
spearman = corr(X, "spearman");
mattrib corr      rowname=varNames colname=varNames
           spearman rowname=varNames colname=varNames;
print "Pearson correlation matrix",
corr;
print "Spearman correlation matrix",
spearman;
quit;
/*

```

OK (1)

```

Se observa como las variables que presentan una mayor correlación son el peso con
las semanas de embarazo, y la
edad del padre y la madre.*/

```

```

/* Ej. 8 */

```

```

* El objetivo que nos planteamos es investigar qué factores pueden estar relacionados
con la prematuridad del parto;
* Hacemos un primer gráfico donde visualizamos las proporciones
prematuro/no-prematuro respecto del número de nacidos;

```

```

title "Proporciones de prematuridad condicionando por número de bebés";
PROC GCHART DATA = prac.dades;
HBAR3D Prematur/ TYPE = PCT GROUP = Numero_nascuts;
RUN;

```

OK (0.75)

No conozco
el efecto de todas
las variables
explicativas.

```

* Nos planteamos ver qué variables están significativamente asociadas con la prematuridad.
Añadiremos las variables que creemos que tiene sentido considerar como
posibles causantes o confusoras
en un modelo de regresión logística:
1- numero de bebés
2- nivel de estudios (según la agrupación que definimos anteriormente)
3- edad de la madre
4- edad del padre
5- nacionalidad española (si/no)
3- sexo del bebé
7- provincia
3- años de relación;

```

solo de
algunas.
Falta decir
que edad del
padre y la
madre parecen
no tener
efecto.
Ver informe
adjunto.

```

PROC LOGISTIC DATA=prac.dades DESCENDING;
CLASS sexe;
CLASS estudis (PARAM=REF REF='ESO/Batxillerat');
CLASS nacionalitat (PARAM=REF REF='Si');
CLASS provincia (PARAM=REF REF='Barcelona');

```

```
MODEL Prematur= sexe estudis edat_mare edat_pare nacionalitat numero_nascuts
    provincia anys_relacio;
RUN;
```

/* Vemos que las variables que aparecen como significativas en el modelo son:
El número de bebés es la variable más asociada.
También vemos que aparecen como significativas: nivel de estudios, nacionalidad española.
*/

/* Ejercicio 9 */
/* Como extra, nos ha interesado investigar las diferencias de edad entre padres y madres.
Puesto que
hemos visto suficiente normalidad en los datos de estas dos variables y son datos
aparejados, realizamos un test t de Student
para datos aparejados.

El objetivo es descubrir si las diferencias de edad entre madres y padres son
significativas y en qué dirección.*/

```
ods graphics on;
proc ttest;
    paired Edat_mare*Edat_pare;
run;
ods graphics off;
```

ok (1)

* En los resultados vemos una ligera diferencia significativa indicando que la
edad media de las madres es menor que la de los padres
en aproximadamente unos 3 años;

/* Ejercicio 10*/

```
* Macro que realiza 3 tipos de gráfico cruzando dos de las variables cualesquiera
en la base de datos teniendo en cuenta si son
categóricas o continuas.;
* type = tipo de gráfico (puede ser: contbycat, contbycont o catbycat);
* var1 = variable en el eje x;
* var2 = variable en el eje y;
%MACRO CompareVars(type,var1,var2);
ods html;
%if &type = contbycat %then %do;
proc sgplot data=prac.dades;
vbox &var1 / group=&var2 ;
run;
%end;
%else %if &type=contbycont %then %do;
proc gplot data = prac.dades;
plot &var1*&var2;
run;
%end;
%else %if &type=catbycat %then %do;
PROC GCHART DATA = prac.dades;
HBAR3D &var1/ TYPE = PCT GROUP = &var2;
RUN;
%end;
ods html close;
%MEND;
```

ok (1)

```
* Reproducimos el gráfico anterior para investigar la relación entre la
prematuridad y el número de nacidos*/
;CompareVars(catbycat,premat,numero_nascuts)

Otros ejemplos de cómo usaríamos la macro;
;CompareVars(contbycat,pes,sexe)
;CompareVars(contbycont,edat_mare,pes)

ods pdf close;
```