

COMPUTACIÓN EN ESTADÍSTICA Y OPTIMIZACIÓN

SAS – EJERCICIOS

Ania Alay

Máster en estadística e investigación operativa

Año académico 2015 – 2016

SESIÓN 2 – GESTIÓN DE BASES DE DATOS

Ejercicio 1 - Importación

Crear la base d:\CURSAS14_15\S02\Ejercicios\GIRONA.SAS7BDAT a partir de los datos de población de la hoja Gerona en el archivo d:\CURSAS14_15\S02\dades_s02\POBMUNCAT_CEN01.XLS. Guardar la sintaxis en el archivo Ej1_S02.SAS.

a) ¿Cuántas observaciones debería tener GIRONA.SAS7BDAT?

221

b) ¿Cuántas observaciones tiene?

457

c) Modifica la sintaxis del fichero Ej1_S02.SAS.

libname e 'ruta';

PROC IMPORT OUT= E.GIRONA

DATAFILE= "ruta\pobmuncat_cen01.xls"

DBMS=EXCEL REPLACE;

RANGE="Girona\$A1:C222";

GETNAMES=YES;

MIXED=NO;

SCANTEXT=YES;

USEDATE=YES;

SCANTIME=YES;

RUN;

EJERCICIO 2 - IMPORTACIÓN

a) A partir de la información que se encuentra en la hoja BCN del fichero POBMUNCAT_CEN01.XLS crear la base BCN.SAS7BDAT con los datos relativos a la población de los diferentes municipios de la provincia de BCN. Guardar la sintaxis correspondiente en el fichero Ej2_S02.SAS. Compruebe que el número de observaciones del fichero BCN.SAS7BDAT es 311. En caso contrario, modifica la sintaxis adecuadamente.

PROC IMPORT OUT= lib.bcn

DATAFILE= "e:\SAS\DadesT2\pobmuncat_cen01.xls"

DBMS=EXCEL REPLACE;

RANGE="bcn\$9:319";

GETNAMES=NO;

MIXED=NO;

SCANTEXT=YES;

USEDATE=YES;

SCANTIME=YES;

RUN;

b) Crea la base d:\CURSAS14_15\S02\Ejercicios\BCN3.SAS7BDAT a partir de BCN.SAS7BDAT sólo con 3 variables con los nombres P1, P2 y P3 y las etiquetas:

Nombre de variable	Etiqueta
P1	Código municipio
P2	Nombre municipio
P3	Población total

data lib.bcn3 (rename = (F1=P1 F3=P2 F5=P3));

set lib.bcn (keep = F1 F3 F5);

label F1='Código municipio' F3='Nombre municipio' F5='Población total';

run;

EJERCICIO 3 - IMPORTACIÓN

a) Crear el fichero d:\CURSAS14_15\S02\Ejercicios\DONES.SAS7BDAT con los datos de la tabla MUJERES que se encuentran en fichero d:\CURSAS14_15\S02\dades_s02\NAC_MARE.MDB. Guardar la sintaxis en Ej3_S03DONES.SAS.

```
PROC IMPORT OUT= e.dones
    DATATABLE= "MUJERES"
    DBMS=ACCESS REPLACE;
    DATABASE="ruta\Nac_mare.mdb";
    SCANMEMO=YES;
    USEDATE=NO;
    SCANTIME=YES;
RUN;
```

b) Modificar la sintaxis de DONES.SAS con el fin de crear HOMES.SAS7BDAT con los datos de la tabla HOMES. Guardar la sintaxis en Ej3_S03HOMES.SAS.

```
PROC IMPORT OUT= e.homes
    DATATABLE= "VARONES"
    DBMS=ACCESS REPLACE;
    DATABASE="ruta\Nac_mare.mdb";
    SCANMEMO=YES;
    USEDATE=NO;
    SCANTIME=YES;
RUN;
```

EJERCICIO 4 - EXPORTACIÓN

El archivo d:\CURSAS14_15\S02\Ejercicios\GIRONA.SAS7BDAT contiene información sobre la población en los diferentes municipios de Girona. El ejercicio consiste en la creación de diferentes bases de datos:

a) Crear el archivo d:\CURSAS14_15\S02\Ejercicios\GIRO_COMES.DAT delimitado por comas.

```
PROC EXPORT DATA= E.Girona
    OUTFILE= "ruta\giro_comes.csv"
    DBMS=CSV REPLACE;
    PUTNAMES=YES;
RUN;
```

b) Crear el archivo d:\CURSAS14_15\S02\Ejercicios\GIRO_TAB.DAT delimitado por tabuladores.

```
PROC EXPORT DATA= E.Girona
    OUTFILE= "d:\giro_TAB.dat"
    DBMS=TAB REPLACE;
    PUTNAMES=YES;
RUN;
```

c) Crear el archivo d:\CURSAS14_15\S02\Ejercicios\GIRO.DAT delimitado por espacios en blanco solo con los municipios que tienen más de 2200 habitantes.

*/*primera manera de hacerlo*/*

```
PROC EXPORT DATA= WORK.Girona (where = (total>2200))
    OUTFILE= "ruta\giro.dat"
    DBMS=DLM REPLACE;
    DELIMITER='00'x;
    PUTNAMES=YES;
RUN;
```

*/*segunda manera*/*

```
data e.giro;
set e.girona;
if total>2200;
```

```
run;  
PROC EXPORT DATA= E.Giro  
    OUTFILE= "d:\giro.dat"  
    DBMS=DLM REPLACE;  
    DELIMITER='00'x;  
    PUTNAMES=YES;
```

```
RUN;
```

d) Crear el archivo d:\CURSAS14_15\S02\Ejercicios\GIRO.XLS asignando el nombre Dades_Girona a la hoja de datos. Guardar la sintaxis en Ej4_S03.SAS.

```
PROC EXPORT DATA= E.Girona  
    OUTFILE= "ruta\giro.XLS"  
    DBMS=EXCEL REPLACE;  
    SHEET="dades_girona";  
RUN;
```

SESIÓN 3 – GESTIÓN DE BASES DE DATOS

Ejercicio 1 – Gestión de archivos. Opciones de Datasets.

El archivo D:\CURSAS15_16\S03\dades_s03\EBPF.SAS7BDAT contiene información sobre todas las Comunidades Autónomas. El ejercicio consiste en crear diferentes archivos a partir de la información que encontrareis en este fichero.

a) Definir un único BLOC DATA que permita crear los siguientes archivos: EBNOR.SAS7BDAT y EBRES.SAS7BDAT. El primero ha de tener todas las variables pero sólo las observaciones del NORD. El segundo ha de tener todas las observaciones excepto las del NORD.

```
data e.ebnor (where = (codi='N')) e.ebres (where = (codi^='N'));  
set e.ebpf;  
run;
```

b) Definir un BLOC DATA que permita crear el archivo EB_613.SAS7BDAT sólo con las observaciones de la 6 a la 13. Guardar el script como Ej1_S03.SAS.

```
data e.eb613;  
set e.ebpf (firstobs=6 obs=13);  
run;
```

Ejercicio 2 – ODS.

Obtener la distribución de frecuencias de las variables *genere*, *escriu* y *esport* y un descriptivo de *edad*, *pes* y *altura* del archivo ESTU1.SAS7BDAT. Guardar los resultados en el archivo EJ2.rtf, añadiendo un título adecuado. Guardar el script como Ej2_S03.SAS.

```
ods rtf file='RUTA\ej2.rtf';  
title "Distribucion de frecuencias";  
options nocenter;  
proc freq data=e.estu1; tables x1 x9 x10; run;  
proc means data=e.estu1; var x2 x3 x4; run;  
ods rtf close;  
title;
```

Ejercicio 3 – Lectura de archivos ASCII.

a) El archivo ESTU96.DAT contiene 10 variables. Se trata de crear la base ESTU96.SAS7BDAT con toda la información del archivo ESTU96.DAT. Utilizar como nombres de variables: X1, X2,..., X10.

```
data e.estu96;  
infile 'RUTA\estu96.dat';  
input x1 $ x2-x9 x10 $;  
run;
```

b) El archivo MATRIMONIS.XLS contiene información sobre los matrimonios efectuados durante los años 1975-2009 en las distintas provincias españolas. Importa la base de datos creando MATRIMONIS.SAS7BDAT.

```
PROC IMPORT OUT= e.matrimonis  
    DATAFILE= "RUTA\matrimonios.xlsx"  
    DBMS=EXCEL REPLACE;  
    RANGE="matrimonis$";  
    GETNAMES=YES;  
    MIXED=NO;  
    SCANTEXT=YES;  
    USEDATE=YES;  
    SCANTIME=YES;  
RUN;
```

c) A partir de MATRIMONIS.SAS7BDAT se trata de crear tres nuevas bases de datos con la información correspondiente a las comunidades autónomas de Cataluña, Castilla la Mancha y Galicia, con nombres CATAL, CASTMAN Y GALIC. Utiliza un único bloque data para crear las tres bases de datos. Guardar el script como Ej3_S03.SAS.

```
data e.catal (keep = Barcelona Girona Lleida Tarragona) e.castman (keep = Albacete Ciudad_Real  
Cuenca Guadalajara Toledo) e.galic (keep = Coruna__a_ Lugo Ourense Pontevedra);  
set e.matrimonis;  
run;
```

Ejercicio 4 – Lectura con informats (I).

a) A partir del archivo D:\CURSAS15_16\S03\dades_s03\HOSP.DAT diseñar un BLOC DATA que permita:

1. Crear la base D:\CURSAS15_16\S03\Ejercicios\HOSP.SAS7BDAT con los siguientes nombres de variables y etiquetas:

NOMBRE	ETIQUETA
D1 (carácter)	Historia del paciente
D2	Fecha de admisión (m/d/y)
D3	Fecha de alta (m/d/y)
D4	Coste de hospitalización
D5	Tipos de diagnóstico

2. Asignar un formato conveniente a las variables D2 y D3. Definir las variables: DIAS = Número de días de hospitalización y COST_D = Gasto medio por día de hospitalización de cada paciente.

```
data e.hosp;  
infile 'RUTA\hosp.dat';  
input d1 $ d2 :mmddyy8. d3 :mmddyy8. d4 d5;  
format d2 d3 ddmmyy10.;  
dias=d3-d2;  
cost_d=d4/dias;  
label dias = 'Dias de hospitalizacion' cost_d='Gasto medio por dia';  
run;
```

b) ¿Cuál es el número medio de días de hospitalización en general y según el tipo de diagnóstico?

```
proc means data=e.hosp; var dias; class d5; run;
```

c) Lista sólo los valores de las variables D1, D2, D3 y DIAS para los que hayan tenido un valor de COST_D superior 800. Guardar este listado en PACIENTS.pdf. Guardar el script como Ej4_S03.SAS.

```
ods pdf file='RUTA\pacientes.pdf';  
proc print data=e.hosp (where = (cost_d>800)); var d1-d3 dias; run;  
ods pdf close;
```

Ejercicio 5 – Lectura con informats (II).

A partir del archivo D:\CURSAS15_16\S03\dades_s03\ASEG.DAT:

a) Crear la base D:\CURSAS15_16\S03\Ejercicios\ASEG.SAS7BDAT con los siguientes nombres de variables y etiquetas. y definir una variable con nombre ANTIG que recoja la antigüedad de los asegurados en años. Puedes utilizar la función TODAY().

NOMBRE	ETIQUETA	FORMATO
V1 (Carácter)	Número póliza	
V2	Fecha alta	dd/mm/yy
V3	Nombre	
V4	Primer apellido	
V5	Segundo apellido	
V6	Fecha nacimiento	dd/mm/yy
V7	Género	

```
data e.ASEG;  
infile 'RUTA\ASEG.dat';  
input v1 $ v2 :ddmmyy8. @15 v3 $12. (v4 v5) ($) @60 v6 ddmmyy10. v7 $;  
format v2 v6 ddmmyy10.;  
antig=(today()-v2)/365;  
run;
```

b) Listar los valores de todas las variables al fichero de resultados.

```
proc print data=e.ASEG; run;
```

c) Listar sólo los valores de las variables V1, V2, V6 y ANTIG al fichero de resultados de forma que las variables V2 y V6 tengan un formato de mes/día/año y que el año se muestre con 4 dígitos. Guardar el script como Ej5_S03.SAS.

```
proc print data=e.ASEG; var v1 v2 v6 antig; format v2 v6 mmddyy10.; run;
```

SESIÓN 4 – GESTIÓN DE BASES DE DATOS

Ejercicio 1. Listados

A partir del archivo MEDI.SAS7BDAT se pide:

a) Listar en la ventana OUTPUT los nombres de las variables, tipo, formato, etiquetas y orden en que se encuentran en el archivo D:\CURSAS15_16\S04\dades_s04\MEDI.SAS7BDAT.

```
ods listing;
```

```
proc contents data=mm.medi;
```

```
run;
```

```
ods listing close;
```

b) Hacer un listado solo con las cuatro primeras variables identificadas por sus etiquetas y sin visualizar el número de observación. Guardar el listado como D:\CURSAS15_16\S04\Ejercicios\LISTADO1.PDF.

```
ods pdf file='ruta\LISTADO1.pdf';
```

```
proc print data=mm.medi noobs label;
```

```
var num_pac data med chol;
```

```
run;
```

```
ods pdf close;
```

c) Utilizando únicamente un PROC PRINT haz un listado en la ventana OUTPUT de todas las variables del archivo identificadas por sus etiquetas y sin visualizar el número de observación, pero sólo para los individuos con Código Médico igual a D. Guardar la sintaxis como Ej1_S04.SAS.

```
ods listing;
```

```
proc print data=mm.medi noobs label;
```

```
where med='D';
```

```
run;
```

```
ods listing close;
```

Ejercicio 2. Listados y ordenaciones.

El archivo D:\CURSAS15_16\S04\dades_s04\BAS11.SAS7BDAT contiene información relativa al número de minutos de conversación telefónica (X) por parte de los trabajadores de dos departamentos (Z) en una empresa. Se pide:

a) Crear la base de datos D:\CURSAS15_16\S04\Ejercicios\BAS11ORD.SAS7BDAT ordenada según departamento y listarla sin mostrar el número de observación.

```
ods listing;
```

```
proc sort data=mm.bas11 out=bas11ord; by z; run;
```

```
proc print data=bas11ord noobs; run;
```

b) Listar la base de datos ordenada según departamento en sentido ascendente y según duración de la llamada en sentido descendente, sin mostrar el número de observación.

```
proc sort data=mm.bas11 out=bas11ord2; by z descending x; run;
```

```
proc print data=bas11ord2 noobs; run;
```

c) Hacer un listado con la duración de cada una de las llamadas efectuadas en cada departamento por separado donde al final aparezca también el tiempo total empleado por cada departamento, así como el tiempo total de todas las llamadas. Guardar el script como Ej2_S04.SAS.

```
proc print data=bas11ord;
```

```
var x;
```

```
by z;
```

```
sum x;
```

```
run;
```

Ejercicio 3. PROC FORMAT (I)

El archivo D:\CURSAS15_16\S04\dades_s04\ENQASSEG.SAS7BDAT contiene las siguientes variables sobre la opinión de unos asegurados de una compañía:

Variable	Descripción
X1	Número encuesta
X2	Año de nacimiento
X3	Género (1= Hombre 2 = Mujer)
X4	Estado civil (1 = Viudo/a, 2 = Separado/Divorciado, 3 = Casado/a y 4 = Soltero/a)
X5	Calidad del servicio de atención al cliente
X6	Calidad del servicio de tramitación de siniestros
X7	Calidad del servicio de reclamaciones

Las variables x5, x6 y x7 recogen la información respecto a la opinión de los asegurados, en una escala de 1 a 5, donde: 1 = Muy buena, 2 = Buena, 3 = Regular, 4 = Mala y 5 = Muy mala, 6 o más = No respuesta.

Se pide:

a) Crea el archivo D:\CURSAS15_16\S05\Ejercicios\Ejer_1\ENQASSEG2.SAS7BDAT a partir del anterior en el que se estén etiquetadas las variables del archivo con su descripción y también los valores de las variables. Guarda el catálogo de formatos en la carpeta D:\CURSAS15_16\S05\Ejercicios\Ejer_3\.

```

ibname w 'ruta\s04\dades_s04';
libname w1 'ruta\s04\dades_s04\ejercicios\Ejer_1';
proc format library = w1;
value opi
1 = 'Muy buena'
2 = 'Buena'
3 = 'Regular'
4 = 'Mala'
5 = 'Muy mala'
6-high = 'NS/NC';
value gene 1='Hombre' 2= 'Mujer';
value estadc 1='Viudo/a' 2='Separado/Div' 3 = 'Casado/a' 4='Soltero';
run;
options fmtsearch=(w1);
data w1.enqasseg2;
set w.enqasseg;
format x5-x7 opi. x3 gene. x4 estadc.;
*label AÑADIR VOSOTROS LAS ETIQUETAS VARIABLE;
run;

```

b) Cerrar la sesión de SAS c) Volver a entrar en SAS

d) Obtener la distribución de frecuencias de las variables x5, x6 y x7. Comprobar que la no respuesta es 3.33%, 2% y 0.67% respectivamente. Guardar la sintaxis como Ej3_S04.SAS.

```

libname w 'ruta\s04\dades_s04';
libname w1 'ruta\s04\dades_s04\ejercicios\Ejer_1';
options fmtsearch=(w1);
proc freq data=w1.enqasseg2;
tables x5 x6 x7;
run;

```

Ejercicio 4. PROC FORMAT (II).

a) A partir del archivo D:\CURSAS15_16\S04\dades_s04\MON00D.SAS7BDAT obtener la distribución de frecuencias de la variable POB según la siguiente agrupación: hasta 14, (14,25], (25,40] y más de 40. Guarda el catálogo de formatos en D:\CURSAS15_16\S04\Ejercicios\Ejer_4\.

```

Libname w2 'ruta\s04\dades_s04\ejercicios\Ejer_2';
proc format library=w2;
value pobl low-14 = 'Hasta 14' 14<-25= '(14,25]' 25<-40= '(25,40]' 40<-high='mas de 40';
run;

```

b) Comprobar que se obtienen los siguientes resultados:

Variable agrupada	Frecuencia
Hasta 14	10
(14,25]	21
(25,40]	10
Más de 40	20

```
options fmtsearch=(w2);  
proc freq data=mon00d2; format pob pobl.; tables pob; run;
```

Ejercicio 5. PROC FORMAT (III).

a) Hacer un PROC CONTENTS del archivo D:\CURSAS15_16\S04\dades_s04\BAS08.SAS7BDAT; ¿qué información importante se obtiene?

```
proc contents data=w.bas08; run;  
proc format library = w;  
value SN
```

```
1 = 'SI'
```

```
2 = 'NO';
```

```
run;
```

/*es posible que aqui os dé error si el archivo FORMATS.SAS7BCAT está protegido contra escritura. En tal caso, poniendo el ratón encima del archivo FORMATS.SAS7BCAT y clicando el botón derecho, id a propiedades y desactivar la opción de solo lectura.*/

b) Asignar las etiquetas 1 = SI, 2 = NO a los valores de V3 a V5, incorporando estas etiquetas en el archivo FORMATS.SAS7BCAT y obtener la distribución de frecuencias de las variables V3 a V5. Guardar la sintaxis como Ej5_S04.SAS.

```
options fmtsearch=(w);  
proc freq data=w.bas08; tables v3-v5; run;
```

Ejercicio 6. PROC FORMAT (IV).

El archivo POL.SAS7BDAT contiene información relativa a los asegurados de una compañía. Queremos obtener dos listados. En ambos solo han de aparecer las variables X1 y X4 pero uno de los listados lo queremos enviar a Barcelona con las siguientes etiquetas para las variables X1 y X4: X1 d=Dona h=Home X4 1=Si 2=No y el otro lo enviamos a Londres con estas etiquetas: X1 d=Female h=Male X4 1=Yes 2=No.

Guarda el catálogo de formatos en \Ejer_6\. Guardar la sintaxis como Ej6_S04.SAS.

```
libname w4 'rutals04\dades_s04\ejercicios\Ejer_4';
```

```
proc format library=w4;  
value $gene 'd'='Dona' 'h'='Home';  
value respc 1 = 'SI' 2 = 'NO';  
value $gender 'd'='Female' 'h'='Male';  
value respa 1 = 'Yes' 2 = 'No';  
run;
```

```
options fmtsearch=(w4);
```

```
proc print data=w.pol; format x1 $gene. x4 respc.;  
var x1 x4;  
run;
```

```
proc print data=w.pol; format x1 $gender. x4 respa.;  
var x1 x4;  
run;
```

SESIÓN 5 – GESTIÓN DE BASES DE DATOS

Ejercicio 1. UNIÓN DE ARCHIVOS (I)

El archivo D:\CURSAS15_16\S05\dades_s05\ASSEG_JOVES.SAS7BDAT contiene una serie de variables sobre los tomadores jóvenes (menores de 30 años) de seguros de salud de una determinada compañía. Del mismo modo, el archivo D:\CURSAS15_16\S05\dades_s05\ASSEG_SENIOR.SAS7BDAT contiene información sobre los tomadores senior (mayores de 50 años) de seguros de salud de la misma compañía. Se pide:

a) Fusiona los archivos en uno solo con nombre D:\CURSAS15_16\S05\Ejercicios\ASSEG_TOT.SAS7BDAT creando una nueva variable que indique la procedencia de la observación.

```
data w.asseg_tot; set w.asseg_joves (in=v1) w.asseg_senior; jove=v1; run;
```

b) Ordena la base de datos D:\CURSAS15_16\S05\Ejercicios\ASSEG_TOT.SAS7BDAT por el número de póliza.

```
proc sort data=w.asseg_tot; by npol; run;
```

c) Calcula la edad media del total de asegurados utilizando PROC MEANS. Guardar la sintaxis como Ej1_S05.SAS.

```
proc means data=w.asseg_tot; var edat; run;
```

Ejercicio 2. UNIÓN DE ARCHIVOS (II)

El archivo ASSEG_SENIOR2.SAS7BDAT contiene información sobre el coste total de las visitas a los servicios de salud de los asegurados senior del archivo ASSEG_SENIOR.SAS7BDAT. Se pide:

a) Fusiona los archivos ASSEG_SENIOR.SAS7BDAT y ASSEG_SENIOR2.SAS7BDAT creando el archivo D:\CURSAS15_16\S05\Ejercicios\ASSEG_SENIORT.SAS7BDAT.

```
proc sort data=w.asseg_senior out=asseg1; by npol; run;
```

```
proc sort data=w.asseg_senior2 out=asseg2; by npol; run;
```

```
data w.asseg_seniort; merge aseg1 aseg2; by npol; run;
```

b) Crea la base de datos D:\CURSAS15_16\S05\Ejercicios\ASSEG_SENCOST.SAS7BDAT sólo con los asegurados que han realizado alguna visita médica (*Nvisitas*) e incluyendo solo las variables número de póliza (*Npol*) y coste medio de las visitas efectuadas por dichos pacientes.

```
data w.asseg_sencost; set w.asseg_seniort;
```

```
if nvisitas>0;
```

```
keep npol coste;
```

```
run;
```

c) Haz un listado del contenido del archivo ASSEG_SENCOST.SAS7BDAT. Guardar la sintaxis como Ej2_S05.SAS.

```
proc print data=w.asseg_sencost;
```

```
run;
```

SESIÓN 6 – REPASO

El archivo SEGUROS.SAS7BDAT contiene 34530 observaciones y variables.

Ejercicio 1. ETIQUETAR.

Etiqueta adecuadamente las variables del archivo y sus valores teniendo en cuenta la información proporcionada en la Tabla 1. Guarda el archivo etiquetado con nombre SEGUROSE.SAS7BDAT y añade etiquetas de valor a las variables SEXO y TIPOP incorporándolas al archivo FORMATS.SAS7BCAT que has descomprimido.

```
libname pp 'ruta';  
proc contents data=pp.seguros; run;  
options fmtsearch=(pp);  
proc format library=pp fmtlib;  
value $sex 'M'='Mujer' 'H'='Hombre';  
value tippol 1='Terceros' 2='Terceros más complementos' 3='Todo riesgo con franquicia' 4='Todo  
riesgo sin franquicia';  
run;
```

```
data pp.segurose; set pp.seguros; format sexo sex. tipop tippol.;  
run;
```

Ejercicio 2. ODS.

Haz un breve análisis descriptivo de las variables SEXO, EDAD, PROVIN, TIPOP PRIMAP y NSINI (utilizando PROC MEANS/PROC FREQ dependiendo del caso). Guarda los resultados en formato RTF, con nombre INFORME.RTF. El informe tiene que tener un título adecuado, estar justificado a la izquierda, no debe aparecer el nombre del procedimiento de SAS empleado, ni la fecha, ni el número de página.

```
ods rtf file='ruta\informe.rtf';  
options nodate nonumber nocenter;  
ods noproctitle;  
proc freq data=pp.segurose; tables sexo provin; run;  
proc means data=pp.segurose; var edad tipop primap nsini; run;  
ods rtf close;
```

Ejercicio 3. LISTADOS.

a) Cuál es el número de póliza correspondiente al asegurado que más ha pagado por su seguro?

```
proc sort data=pp.segurose out=segord; by descending primap; run;  
proc print data=segord (obs=1); run;
```

b) Haz un listado del número de póliza y de la última prima pagada de los asegurados que residen en la provincia de Barcelona, tienen más de 60 años y tienen contratada una póliza a todo riesgo con franquicia. Cuántas observaciones has listado? Cuál es el importe total recaudado por la compañía en concepto de primas correspondiente a este grupo de asegurados?

```
proc print data=pp.segurose (where = (provin=1 & edad>60 & tipop = 3)) n;  
var npoliza primap;  
sum primap;  
run;
```

Ejercicio 4. GESTIÓN BASES DATOS / EXPORTACIÓN.

Crea un archivo de EXCEL que se llamará SEGUROSM.XLSX que contendrá sólo las observaciones correspondientes mujeres de Girona menores de 50 años y que no hayan declarado ningún siniestro el último año. El archivo solo ha de contener las variables NPOLIZA, EDAD, TIPOP y PRIMAP que aparecerán en la

primera línea de la hoja de cálculo renombradas como X1, X2, X3 y X4. Cuántas observaciones tiene SEGUROSM.XLSX?

```
data pp.seguosm (keep = npoliza edad tipop primap rename = (npoliza = x1 edad=x2 tipop=x3 primap=x4));  
set pp.seguose (where = (provin=3 & sexo='M' & edad<50 & nsini=0));  
run;
```

```
PROC EXPORT DATA= pp.SegurosM  
    OUTFILE= "ruta\segurom.xls"  
    DBMS=EXCEL REPLACE;  
    SHEET="seg";  
RUN;
```

Ejercicio 5 – IMPORTACIÓN / FUSIÓN.

Importa la base de datos SINIESTROS.TXT a SAS creando la base de datos SINIESTROS.SAS7BDAT.

La base de datos creada, contiene información sobre las siguientes variables medidas sólo para los que han declarado algún siniestro:

VARIABLE	DESCRIPCIÓN
NPOLIZA	Número de póliza del asegurado
COSTESIN	Coste total de los siniestros declarados en el año
OPINION1	Valoración de la tramitación del siniestro (de 0 a 10)
OPINION2	Valoración de la rapidez en la prestación de cobertura (de 0 a 10)
OPINION3	Valoración de la compañía de seguros en general (de 0 a 10)

a) Fusiona la base de datos SEGUROSE.SAS7BDAT con SINIESTROS.SAS7BDAT creando SEGSINI.SAS7BDAT quedándote sólo con los individuos que han declarado algún siniestro.

```
PROC IMPORT OUT= pp.Siniestros  
    DATAFILE= "ruta\siniestros.txt"  
    DBMS=DLM REPLACE;  
    DELIMITER=' ';  
    GETNAMES=YES;  
    DATAROW=2;
```

```
RUN;
```

b) Crea una variable nueva que se llamará COSTEMED que recoja para cada individuo el coste medio de los siniestros declarados durante el último año.

c) Haz un listado con los números de póliza de los cinco asegurados que han tenido un coste medio por siniestro más elevado.

d) Crea una variable nueva que se llamará OPIMED que sea la media de las valoraciones otorgadas en las variables de opinión (OPINION1, OPINION2 y OPINION3).

e) ¿Qué porcentaje de los asegurados que han declarado algún siniestro han otorgado una valoración media a las variables de opinión superior a 5?

```
proc sort data=pp.seguose; by npoliza; run;  
proc sort data=pp.siniestros; by npoliza; run;  
data pp.segsini; merge pp.seguose (in=v1) pp.siniestros (in=v2); by npoliza; if v1=v2; run;  
data pp.segsini; set pp.segsini;  
opinamed=mean(opinion1,opinion2,opinion3);  
costemed=costesin/nsini;  
opicat=(opinamed>5);  
run;  
proc sort data=pp.segsini; by descending costemed; run;  
proc print data=pp.segsini (obs=5) ; var npoliza; run;  
proc freq data=pp.segsini; tables opicat; run;
```

SESIÓN 8 – CREACIÓN DE VARIABLES Y PROCS

Ejercicio 1. CREACIÓN DE VARIABLES

1. Crea una base temporal en D:\CURSAS15_16\S08\dades_s08\Ejercicios\ con nombre ALEA.SAS7BDAT que contenga 350 observaciones aleatorias de 5 variables: $X1 \sim N(0,1)$, $X2 \sim N(10,2)$, $X3 \sim U(0,1)$, $X4 \sim U(0,10)$ y $X5 \sim U(-10,10)$ utilizando las funciones RANUNI y RANNOR con un valor de semilla de 954367. La base de datos resultante, sólo ha de tener esas 5 variables.

DATA alea;

do i = 1 to 350;

x1=rannor(954367);

x2=rannor(954367)*sqrt(2)+10;

x3=ranuni(954367);

x4=ranuni(954367)*10;

x5=-10+ranuni(954367)*20;

output;

end;

drop i;

run;

proc print data= alea;

run;

proc means data=alea; run;

2. Crea a partir de ALEA.SAS7BDAT otra base de datos que se llamará ALEA2.SAS7BDAT con las siguientes variables:

- Y1: valor absoluto de X1
- Y2: signo de X1
- Y3: raíz cuadrada de X4
- Y4: redondeo de X2 a un decimal
- Y5: exponencial de X1
- Y6: máximo de X3, X4 y X5
- Y7: mínimo de X3, X4 y X5
- Y8: media de los valores absolutos de X1 y X5.

Guardar la sintaxis como Ej1_S08.SAS.

DATA alea2;

set alea;

y1=abs(x1);

y2=sign(x1);

y3=sqrt(x4);

y4=round(x2,0.1);

y5=exp(x1);

y6=max(x3,x4,x5);

y7=min(x3,x4,x5);

y8=mean(abs(x1),abs(x5));

drop x1-x5;

run;

proc means data=alea2; run;

Ejercicio 2. CREACIÓN DE VARIABLES Y PROCEDIMIENTOS

El archivo D:\CURSAS15_16\S08\dades_s08\FAMI.SAS7BDAT presenta el porcentaje de renta destinada a distintos tipos de gasto por diferentes individuos: gasto en ocio, comida, ropa y vivienda. Se pide:

1. ¿Cuál es el gasto medio de los individuos en vivienda?
2. ¿Cuál es el valor máximo del gasto en ocio?
3. ¿Cuál es el valor mínimo del gasto en comida?

```
libname w 'ruta';  
libname ej 'ruta\ejercicios';  
proc means data=w.fami; run;
```

4. Crea la base de datos D:\CURSAS15_16\S08\dades_s08\Ejercicios\FAMI2.SAS7BDAT con las siguientes variables:

- C0 que valdrá 1 para los individuos incongruentes (con gasto total superior al 100%).
- C1 que valdrá 1 para los individuos que destinan como mínimo un 50% de su renta a gastos y 0 en caso contrario.
- C2 que valdrá 1 para los individuos que tengan un gasto en vivienda superior a su gasto en los otros tres conceptos y 0 en caso contrario.
- C3 que valdrá 1 para los individuos que tengan un gasto en comida más ropa superior al gasto en vivienda y 2 en caso contrario.
- C4 que valdrá SUP70 para los que su gasto en ropa y comida sea como mínimo el 70% de su gasto en vivienda. En caso contrario, deberá valer INF70.

```
data ej.fami2; set w.fami;  
if oci+menjar+roba+habitatge>100 then c0=1;  
if oci+menjar+roba+habitatge>50 then c1=1;  
else c1=0;  
if habitatge>oci+menjar+roba then c2=1;  
else c2=0;  
if habitatge<menjar+roba then c3=1;  
else c3=2;  
if (roba+menjar>0.70*habitatge) then c4='sup70';  
else c4='inf70';  
run;
```

5. ¿Qué porcentaje de individuos incongruentes hay? **Ver C0 = 1**

6. ¿Qué porcentaje de individuos tienen un gasto en vivienda superior a los otros tres conceptos?

```
proc freq data=ej.fami2; tables c0 c2 /missing; run;
```

7. ¿Qué porcentaje de los que gastan más de un 25% en vivienda gastan menos de un 15% en ocio?
Guardar la sintaxis como Ej2_S08.SAS.

```
data ej.fami2; set ej.fami2;  
if habitatge>25 then habit=1;  
if oci<15 then oc=1;  
run;  
proc freq data = ej.fami2;  
tables habit * oc / missing;  
run;
```

Ejercicio 3. CREACIÓN DE VARIABLES Y PROCEDIMIENTOS

La base de datos D:\CURSAS15_16\S08\dades_s08\SOCIS.SAS7BDAT contiene información relativa a los socios de una entidad deportiva. A partir de esta base, crear la base de datos D:\CURSAS15_16\S08\dades_s08\Ejercicios\SOCIS2.SAS7BDAT que incluya la nueva variable ANTIC con la antigüedad y la variable EDAT con la edad del socio. Ambas variables se calcularán a día de hoy, utilizando la función TODAY(). La entidad está pensando cambiar las cuotas mensuales de sus socios, de manera que:

- Si el socio tiene menos de 40 años, pagará la cuota básica de 35 Euros/mes.
- Si la edad está entre 40 y 64 (ambos incluidos) la cuota básica se incrementa en un 40%.
- Si la edad es superior a 64 tendrá un 20% de descuento sobre la cuota básica.
- Para los socios de los cuales no conozcamos su edad, queremos que el valor de la cuota sea P, donde P simboliza un dato no disponible o pendiente de determinar.

1. Crea la variable CUOTA.

```
data ej.socis2; set w.socis;  
antic=(today()-data_i)/365.25;  
edat=(today()-data_n)/365.25;  
if edat<40 then cuota=35;  
else if (edat>=40 and edat<=64) then cuota = 1.4*35;  
else if (edat>64) then cuota = 0.8*35;  
if missing(edat)=1 then cuota=.p;  
run;
```

2. ¿Cuál es la cuota media de las mujeres? ¿Y de los hombres?

```
proc means data=ej.socis2; var cuota; class genere; run;
```

3. ¿Qué antigüedad tiene la mujer que más años lleva como socia? ¿y el hombre que más años lleva como socio?

```
proc means data=ej.socis2; var antic; class genere; run;
```

4. Si a partir del próximo mes, todos los socios pasan a pagar la nueva cuota, ¿cuál será el total de ingresos de la entidad en concepto de cuotas el mes próximo?

Guardar la sintaxis como Ej3_S08.SAS.

```
proc means data=ej.socis2 sum; var cuota; run;
```

Ejercicio 4. CREACIÓN DE VARIABLES Y PROCEDIMIENTOS

1. El archivo D:\CURSAS15_16\S08\dades_s08\ENQASSEG.SAS7BDAT contiene las siguientes variables sobre la opinión de unos asegurados de una compañía:

Variable	Descripción
X1	Número encuesta
X2	Año de nacimiento
X3	Género (1= Hombre 2 = Mujer)
X4	Estado civil (1 = Viudo/a, 2 = Separado/Divorciado, 3 = Casado/a y 4 = Soltero/a)
X5	Calidad del servicio de atención al cliente
X6	Calidad del servicio de tramitación de siniestros
X7	Calidad del servicio de reclamaciones

Las variables x5, x6 y x7 recogen la información respecto a la opinión de los asegurados, en una escala de 1 a 5, donde: 1 = Muy buena, 2 = Buena, 3 = Regular, 4 = Mala y 5 = Muy mala, 6 o más = No respuesta. Crea una nueva base de datos D:\CURSAS15_16\S05\dades_s09\Ejercicios\ENQASSEG.SAS7BDAT de manera que:

1. Las variables estén correctamente etiquetadas.
2. La no respuesta en las variables x5, x6 y x7 pase a valer P.

3. ¿Qué porcentaje de no respuesta hay en la variable x5?, ¿y en x6? ¿y en x7?
4. ¿Cuál es la valoración media dada por los asegurados al servicio de atención al cliente? ¿y a la tramitación de siniestros? ¿y al servicio de reclamaciones?
5. Crea dos nuevas variables que se llamarán VMED y VSTDEV que serán respectivamente la valoración media y la desviación típica de las puntuaciones asignadas por cada asegurado a los servicios de atención al cliente, tramitación de siniestros y reclamaciones.

Guardar la sintaxis como Ej4_S08.SAS.

```
data ej.enqasseg; set w.enqasseg;
label X1='Número encuesta'
X2='Año de nacimiento'
X3='Género (1= Hombre 2 = Mujer)'
X4='Estado civil (1 = Viudo/a, 2 = Separado/Divorciado, 3 = Casado/a y 4 = Soltero/a)'
X5='Calidad del servicio de atención al cliente'
X6='Calidad del servicio de tramitación de siniestros'
X7='Calidad del servicio de reclamaciones';
array V(3) x5-x7;
do i=1 to 3;
  if V(i) >=6 then V(i)=.p;
end;
drop i;
run;

proc freq data=ej.enqasseg; tables x5-x7/missing; run;
proc means data=ej.enqasseg; var x5-x7; run;
data ej.enqasseg; set ej.enqasseg; vmed = mean(of x5--x7); vstdev=std(of x5--x7); run;
```

Ejercicio 5. CREACIÓN DE VARIABLES Y PROCEDIMIENTOS

El archivo D:\CURSAS15_16\S08\dades_s08\SINIESTROS.SAS7BDAT contiene las siguientes variables:

Variable	Descripción
NPOLIZA	Número póliza
SEXO	Género (H= Hombre M = Mujer)
N_SIN	Número de siniestros declarados durante el año
COSTE	Coste total de los siniestros declarados durante el año

Se pide:

1. ¿Cuántos siniestros se han declarado en la cartera?
2. ¿Cuál es el coste total de los siniestros declarados en la cartera?

```
proc means data=w.siniestros sum;
var n_sin coste;
run;
```

3. ¿Qué porcentaje de mujeres no ha declarado ningún siniestro? ¿y de hombres?

```
proc freq data=w.siniestros;
tables sexo*n_sin;
run;
```

4. Haz un histograma de la variable COSTE con curva normal superpuesta y calcula los cuartiles.

```
proc univariate data=w.siniestros;
var coste;
histogram /normal;
```

run;

5. Crea una variable que se llamará COSTE1 que recoja el coste medio de cada siniestro declarado por los asegurados. Esta variable sólo se ha de calcular para los individuos que han declarado algún siniestro.

Etiqueta la nueva variable como "Coste medio". Guarda la base de datos como

D:\CURSAS15_16\S09\dades_s09\SINIESTROS2.SAS7BDAT

data sin;

set w.siniestros;

if n_sin>0 then coste1=coste/n_sin;

run;

6. ¿Cuál es el promedio y la varianza del coste medio de los siniestros declarados por los hombres? ¿y por las mujeres? Guardar la sintaxis como Ej5_S08.SAS.

proc means data=sin mean var;

var coste1;

class sexo;

run;

Ejercicio 6. ARRAYS (I)

A partir del archivo DAT1.SAS7BDAT crea la base permanente DAT1.SAS7BDAT que contenga sólo 100 nuevas variables con nombres Y1,...,Y100, una para cada X, de manera que:

- Y = raíz cuadrada del valor absoluto de X si el valor de X es negativo
- Y = raíz cuadrada de X si el valor de X es positivo

Guardar la sintaxis como Ej6_S08.SAS.

DATA ej.dat1 (DROP = COUNT x1--x100);

SET w.dat1;

ARRAY VAL(100) X1 - X100;

ARRAY YY(100) Y1 - Y100;

DO COUNT = 1 TO 100;

YY(COUNT)=sqrt(abs(val(count)));

END;

RUN;

Ejercicio 7. ARRAYS (II)

A partir del archivo DAT2.SAS7BDAT se pide:

a) Crea la base permanente DAT2SOL.7BDAT de manera que el valor -99 de todas las variables pasen a missing de sistema.

DATA ej.dat2sol (DROP = COUNT);

SET w.dat2;

ARRAY XX(100) X1 - X100;

DO COUNT = 1 TO 100;

if (XX(COUNT)=-99) then XX(COUNT)=.;

END;

if (nmiss(of x1--x100)>53) then s=1;

else s=0;

RUN;

b) Crea para cada individuo una variable con nombre S que valga 1 si el número total de observaciones missing es superior a 53 y 0 en caso contrario. ¿qué porcentaje de individuos tienen un número total de variables missings superior a 53?

proc freq data=ej.dat2sol; tables s; run;