

NOM: _____

Els arxius necessaris per fer aquest test els teniu a la carpeta "Test2: Fitxers" que hi ha al campus virtual. Haureu d'entregar només el programa que heu creat per resoldre aquest test a través del campus virtual.

Exercici 1.

La base de dades CARTERA conté informació sobre una cartera de clients d'una companyia d'assegurances. Les variables són:

Poliza	Número de póliza
Sri	Zona de circulación
Sexo	Sexo del asegurado (1 = Hombre, 0 = Mujer)
Gdi	Cobertura contra incendios
Sin	Número de siniestros en 2014
Renueva	Renueva la póliza para 2015
Potencia	Potencia del vehículo
Edad	Edad del asegurado
ant_comp	Antigüedad en la compañía
ant_perm	Antigüedad del permiso de conducir
ant_veh	Antigüedad del vehículo
data_s	Fecha del último siniestro ocurrido en 2014
coste	Coste total de los siniestros ocurridos en 2014
valora1	Valoracion pregunta 1
valora2	Valoracion pregunta 2
valora3	Valoracion pregunta 3
valora4	Valoracion pregunta 4
valora5	Valoracion pregunta 5

1. Etiqueta convenientment els valors de la variable SEXO així com els valors de la variable EDAD segons aquestes tres agrupacions: menor de 30, entre 30 y 50, major de 50. Afegeix els formats al catàleg de formats proporcionat. Quin percentatge d'assegurats són menors de 30 anys?
RESPOSTA: _____ 16.08 _____

```
libname exa 'd:\';  
  
options fmtsearch=(exa);  
  
*1;  
  
proc format library=exa;  
value sexef 1='Hombre' 0='Mujer';  
value edadf low-<30='Menor de 30' 30-50='Entre 30 y 50' 50<-high='Mayor de 50';  
run;  
  
proc freq data=exa.cartera;  
table edad;  
format edad edadf.;  
run;
```

2. Crea una nova variable que es dirà *sinb* que valgui 1 si l'assegurat ha tingut algun sinistre i 0 en cas contrari. Construeix una taula de contingència entre la variable *sinb* i la variable *edad* recodificada segons les agrupacions definides en l'apartat anterior.

Creus que existeix independència entre el grup d'edat i el fet de tenir o no tenir sinistres?
RESPOSTA: ____No es rebutja la independència (p-valor=0.2595)_____

```
data cartera;
set exa.cartera;
sinb=(sin>0);
run;

proc freq data=cartera;
tables sinb*edad / expected chisq;
format edad edadf.;
run;
```

3. Calcula els següents estadístics per a la variable POTENCIA segons les categories de la variable SEXO:

	Mediana	Quartil 3	Asimetria	Rang	Coefficient de Variació
Hombre	70	90	3.718	750	38.36
Mujer	80	95	0.489	320	37.06

```
proc means data=exa.cartera median q3 skew range cv;
var potencia;
class sexo;
format sexo sexef.;
run;
```

4. Crea una base de dades amb format separat per comes que es digui SINISTRES.CSV que contingui només als individus que han tingut algun sinistre. Volem que aquesta base de dades tingui les variables *poliza*, *sexo*, *sin*, *renueva*, *edad* y a més a més una nova variable que se dirà *costem* que sigui el cost mitjà dels sinistres que cada assegurat ha tingut.
Quantes observacions té aquest arxiu? RESPOSTA: ____1224_____

```
data cartera (keep=poliza sexo sin renueva edad costem);
set cartera (where=(sinb=1));
costem=coste/sin;
run;

PROC EXPORT DATA= WORK.CARTERA
OUTFILE= "d:\sinistres.csv"
DBMS=CSV REPLACE;
PUTNAMES=YES;
RUN;
```

5. Quin és el cost total dels sinistres que va haver de pagar la companyia durant 2014 per homes i dones? I en total? Completa aquesta taula:

	Cost total dels sinistres
Hombre	3.823.477
Mujer	1.294.540
Total	5.118.017

Fes un diagrama de barres horitzontal en 3D, on la longitud de cada barra representi el cost total per sexe.

```
proc means data=exa.cartera sum printalltypes;
var coste;
class sexo;
format sexo sexef.;
run;

proc gchart data=exa.cartera;
hbar3d sexo /discrete type=sum sumvar=coste sum;
format sexo sexef.;
run;
```

6. Las variables VALORA1 – VALORA5 contén les respostes dels assegurats a 5 preguntes de valoració sobre la satisfacció amb la companyia en una escala de 0 a 10.
- Quin percentatge dels assegurats ha contestat a totes i cadascuna de les cinc preguntes?
RESPOSTA: ____81.1%____
 - Quin percentatge d'assegurats s'ha deixat al menys dues preguntes per contestar?
RESPOSTA: ____1.54%____

```
data cartera2;
set exa.cartera;
valoraval=nmiss(of valora1--valora5);
ind=(valoraval>=2);
run;

proc freq data=cartera2;
tables valoraval ind;
run;
```

7. Quin percentatge dels assegurats que van tenir sinistres durant 2014 van tenir el seu últim sinistre (variable *data_s*) abans del 15/9/2014? RESPOSTA: ____71.16%____

```
data cartera2;
set cartera2;
if missing(data_s)=0 then data_a=(data_s<mdy(9,15,2014));
run;

proc freq data=cartera2;
tables data_a;
run;
```

8. A partir de les cinc preguntes de valoració (VALORA1, .. VALORA5) crea les següents variables:

VALMED: mediana de VALORA1, ... VALORA5.

VALRG: rang de VALORA1, ... VALORA5.

VALCOEF: coeficient de variació de VALORA1, ... VALORA5.

VALBIN: 1 si la màxima puntuació atorgada a les preguntes VALORA1, ... VALORA5 contestades per l'assegurat ha estat de 8 punts com a mínim, 0 en cas contrari.

Quin dels cinc aspectes valorats en aquestes preguntes ha estat el que en mitjana ha obtingut puntuacions més elevades? RESPOSTA: __El primer (5.01)_____

```
data cartera2;
set cartera2;
valmed=median(of valora1--valora5);
valrg=range(of valora1--valora5);
valcoeff=cv(of valora1--valora5);
valbin=(max(of valora1--valora5)>=8);
run;

proc means data=cartera2;
var valora1--valora5;
run;
```

9. Calcula el coeficient de correlació lineal i les covariàncies entre *edad* i *ant_perm*. Guarda les matrius de correlacions i de covariàncies en una base de dades que es dirà CORREL.SAS7BDAT. Quina conclusió treus sobre la relació entre aquestes variables?

RESPOSTA:_____Correlació positiva estadísticament significativa_____

```
proc corr data=exa.cartera cov out=exa.correl;
var edad ant_perm;
run;
```

10. Fes un diagrama de dispersió entre les variables *potencia* (eix horitzontal) i *coste* (eix vertical) diferenciant entre homes i dones, però només pels que han tingut sinistres i tenen un vehicle de potència inferior a 300. Volem que els punts corresponents als homes apareguin amb un quadrat verd i els punt corresponents a les dones apareguin amb un cercle vermell. Volem com a títol del gràfic "Diagrama de dispersió entre edat i potència del vehicle". A més a més, l'eix vertical el volem de 0 a 50000 amb increments de 5000, i amb 4 marques de graduació menors. L'eix horitzontal el volem de 0 a 300 amb increments de 50, també amb 4 marques de graduació menors. La llegenda del gràfic la volem amunt i al centre.

```
title "Diagrama de dispersió entre edat i potència del vehicle";
symbol1 c=red v=dot;
symbol2 c=green v=square;
axis1 order=(0 to 300 by 50);
axis2 order=(0 to 50000 by 5000);
legend1 position=(top center);
proc gplot data=exa.cartera (where=(potencia<300 & sin>0));
plot coste*potencia=sexo / legend=legend1 haxis=axis1 hminor=4 vaxis=axis2 vminor=4;
format sexo sexef.;
run;
```