**SOLUCION CONTROL 3º GRUPO B1**

1) Identificar las siguientes sentencias con los procedimientos con los que se relacionan:

a) Scatter🡪 PROC G3D

b) bubble🡪PROC GPLOT

c) Exclude🡪 PROC COPY , PROC DATASETS

d) Cdfplot🡪 PROC UNIVARIATE

2) A partir del conjunto de datos ligas y tras submitir un procedimiento Rank se cra el conjunto de datos SAS ranking. Escibir las sentencia necesarias para que se obtenga ese resultado.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | | ligas | | | Liga | X | | ACC | 120 | | B10 | 110 | | B8 | 50 | | P10 | 22 | | EST | 118 | | |  |  |  | | --- | --- | --- | | Ranking | | | | Liga | X | RX | | ACC | 120 | 1 | | B10 | 110 | 1 | | B8 | 50 | 0 | | P10 | 22 | 0 | | EST | 118 | 1 | |

**PROC** **RANK** DATA=LIGAS OUT=RANKING GROUPS=**2**;

VAR X;

RANKS RX;

**RUN**;

3) Dado el conjunto de datos A y el conjunto de datos B, crear el conjunto de datos C

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  | | --- | --- | --- | | **A** | | | | **Comunidad** | **Grupo** | **Destino** | | Andalucia | Zeta | New York | | Andalucia | Beta | Chicago | | Baleares | Riva | Atenas | | Baleares | Zote | Lisboa | | |  |  |  | | --- | --- | --- | | **B** | | | | **Comunidad** | **Origen** | **Precio** | | Andalucia | Jerez | 1000 | | Baleares | Manacor | 5000 | | Galicia | Tuy | 12000 | |

a) Data C; merge A B ; by Comunidad; drop Origen; run;

| **Comunidad** | **Grupo** | **Destino** | **Precio** |
| --- | --- | --- | --- |
| Andalucia | Zeta | New York | 1000 |
| Andalucia | Beta | Chicago | 1000 |
| Baleares | Riva | Atenas | 5000 |
| Baleares | Zote | Lisboa | 5000 |
| Galicia |  |  | 12000 |

b) Data C; set A; by Comunidad; u=substr(Comunidad,3,4); if last.comunidad then output; run;

| **Comunidad** | **Grupo** | **Destino** | **u** |
| --- | --- | --- | --- |
| Andalucia | Beta | Chicago | dalu |
| Baleares | Zote | Lisboa | lear |

4) Simplifica lo máximo que puedas la siguiente sentencia INPUT:

Input v1 1-3 v2 4-5 v3 6-7 v4 8-9 v5 9-10 v6 11-12 v7 13-14 v8 15-16;

INPUT V1 1-3 (V2-V4) (2.) @9 (v5-v8) (2.);

5) Escribir una sentencia que permita calcular la probabilidad de que una binomial de parametros p=0.4 y n=17 sea mayor o igual que 3 y menor que 10.

PROB=CDF(‘BINOMIAL’,9,0.4,17)- CDF(‘BINOMIAL’,2,04,17);

6) Describeme el contenido del conjunto de datos B:

Proc means data=A nway; class year; var km; output out=B mean= mode= / autoname; run;

**B contiene tantas observaciones como valores tenga la variable year. La opción nway hace que no se creen observaciones para las marginales que en este caso sólo seria la media global. Las variables de este conjunto de datos son: year, \_type\_, \_freq\_ mas km\_mean y km\_mode.**

7) Escribe el conjunto de datos PP

Data pp; input x y @; cards;

1 2 3 4 5 6 7

| **x** | **y** |
| --- | --- |
| 1 | 2 |

8) Si tenemos el siguiente conjunto de datos:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  | | --- | --- | --- | | Conjunto de datos A | | | | Sexo | Toma\_vitaminas | Total | | Masculino | Si | 160 | | Masculino | No | 253 | | Femenino | Si | 45 | | Femenino | No | 332 | | a) Escribir las sentencias que creas necesarias para contrastar que la proporción de personas que toman vitaminas no depende del sexo.  b) En una figura representar dos diagramas de tartas (una por cada sexo) con sectores proporcionales al total de personas que toman o no toman vitaminas. (un sector representa a la contestación Si y otro a la contestación No.) |

1. proc freq data=A;

tables sexo\*toma\_vitaminas / chisq;

weight total;

run;

b) proc gchart data=A;

pie toma\_vitaminas / sumvar=total across=2 group=sexo; run;

**SOLUCION CONTROL 3º GRUPO B2**

1) Indicarme lo que hacen las siguientes sentencias:

Pattern v=e c=red; las areas serán transparente y su contorno tendrá color rojo

Symbol i=r w=4 ; sentencia que permite crear una recta de regresión como método de interpolación para una representación de puntos en una grafica del tipo y (eje vertical) frente a x (eje horizontal). La opción w=4 permite que esta linea de regresión lineal tenga una anchura de valor 4 puntos.

Axis2 value=(h=2 f=arial ) order=2 to 10 by 2; Crea un formato de eje donde los valores a representar en él son lo pares del 2 al 10. Estos valores se escribirán con fuente tipo arial y una altura del doble de lo normal.

2) Con que sentencias relacionas las siguientes opciones:

a) Major= Axis

b) down= Pie

c) starmax= Star

d) Bs= Title , Note, footnote

3) Si en el conjunto de datos respiratorios tenemos la tasa de fallecidos por cada 10000 habitantes debido a problemas respiratorios (variable fallecidos) en España por grupo de edad (variable g\_edad) y sexo (variable sexo) en un año (year). Se pide:

a) Escribir el procedimiento que creas necesario para poder representar gráficamente la evolución de la tasa de muerte por año para el grupo de edad ‘de 19-45 años’ (uno de los posibles valores) según el sexo (una figura con dos curvas una para sexo=masculino y otra para sexo=femenino).

Proc gplot data=respiratorios;

where g\_edad=’de 19-45 años’;

Plot fallecidos\*year=sexo; run;

b) Si además tenemos una variable denominada poblacion, que contiene el total poblacional de cada grupo de edad y sexo en ese año . Crear un conjunto de datos que contenga el número total de muertes por año y sexo. (conjunto de datos con solo 3 variables).

**Data respiratorios; set respiratorios; total\_muertes=fallecidos\*poblacion/10000; run;**

**Proc means Data=respiratorios nway;**

**Class year sexo;**

**Var total\_muertes ;**

**Output out=nuevo (drop=\_type\_ \_freq\_) sum=;**

**Run;**

c) Con el conjunto de datos creado en b crear una gráfica en la que se represente la evolución de las muertes a lo largo de los años, con dos curvas una para cada sexo en la misma figura.

proc gplot data=nuevo;

PLOT TOTAL\_MUERTES\*YEAR=SEXO /LEGEND;

RUN;

4) Supongamos que tenemos un fichero externo A.txt (en c:\) del que queremos leer las variables x1, x2 y x3

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| conteniendo 3 observaciones. Escribir las sentencia infile necesaria para que esto se produzca correctamente si los datos no están correctamente columnados | |  |  |  | | --- | --- | --- | | A | | | | 1 | 2 | 3 | | 4 |  |  | | 5 | 6 |  | |

Infile ’c:\A.txt’ missover;

5) Dado el conjunto de datos A y el conjunto de datos B, crear el conjunto de datos C

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  | | --- | --- | --- | | **A** | | | | **Comunidad** | **Grupo** | **Destino** | | Andalucia | Zeta | New York | | Andalucia | Beta | Chicago | | Baleares | Riva | Atenas | | Baleares | Zote | Lisboa | | |  |  |  | | --- | --- | --- | | **B** | | | | **Comunidad** | **Origen** | **Precio** | | Andalucia | Jerez | 1000 | | Baleares | Manacor | 5000 | | Galicia | Tuy | 12000 | |

a) Data C; merge A B ; by Comunidad; drop origen destino; ; run;

| **Comunidad** | **Grupo** | **Precio** |
| --- | --- | --- |
| Andalucia | Zeta | 1000 |
| Andalucia | Beta | 1000 |
| Baleares | Riva | 5000 |
| Baleares | Zote | 5000 |
| Galicia |  | 12000 |

b) Data C; update a (in=a) b; by Comunidad; if a ; run;

| **Comunidad** | **Grupo** | **Destino** | **Origen** | **Precio** |
| --- | --- | --- | --- | --- |
| Andalucia | Zeta | New York | Jerez | 1000 |
| Andalucia | Beta | Chicago |  | . |
| Baleares | Riva | Atenas | Manacor | 5000 |
| Baleares | Zote | Lisboa |  | . |

6) Indicar los errores si es que los hay de las siguientes sentencias:

Array test(20) testbase test0-test18; NO TIENE ERRORES

Array iniciales (5) $1 (1,2,3,4,5); FALTA PONER ENTRE COMILLAS LOS VALORES 1,2,3,4,5 al ser alfanumerica

Array q (10); NO TIENE ERRORES

7) Simplifica lo máximo que puedas la siguiente sentencia INPUT:

Input v1 1-3 v2 4-5 v3 6-7 v4 8-9 v5 9-10 v6 11-12 v7 13-14 v8 15-16;

INPUT V1 1-3 (V2-V4) (2.) @9 (v5-v8) (2.);

**Solucion 3º control A1**

1) Escribir una sentencia que permita calcular la probabilidad de que una binomial de parametros p=0.4 y n=17 sea mayor o igual que 3 y menor que 10.

PROB=CDF(‘BINOMIAL’,9,0.4,17)- CDF(‘BINOMIAL’,2,04,17);

2) Si tenemos el siguiente conjunto de datos:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  | | --- | --- | --- | | Conjunto de datos A | | | | Sexo | Toma\_vitaminas | Total | | Masculino | Si | 160 | | Masculino | No | 253 | | Femenino | Si | 45 | | Femenino | No | 332 | | a) Escribir las sentencias que creas necesarias para contrastar que la proporción de personas que toman vitaminas no depende del sexo.  b) Escribir las sentencias que creas necesarias utilizando el procedimiento transpose para Crear el conjunto de datos B   |  |  |  |  | | --- | --- | --- | --- | | **Cuentas** | **Sexo** | **TomaSi** | **TomaNo** | | Total | Masculino | 160 | 253 | | Total | Femenino | 45 | 332 | |

1. proc freq data=A;

tables sexo\*toma\_vitaminas / chisq;

weight total; run;

b) proc transpose data=a out=B prefix=Toma name=cuentas;

by sexo;

id toma\_vitaminas;

var total;

run;

3) Indicar el contenido (nº de observaciones y nombre de las variables) del conjunto de datos B:

Proc means data=A nway; class year; var km; output out=B mean= mode= / autoname; run;

**B contiene tantas observaciones como valores diferentes tenga la variable year. La opción nway hace que no se creen observaciones para las marginales que en este caso sólo seria la media global. Las variables de este conjunto de datos son: year, \_type\_, \_freq\_ km\_mean y km\_mode.**

4) Escribe el conjunto de datos PP

Data pp; input x y @; cards;

|  |  |
| --- | --- |
| **X** | **Y** |
| **1** | **2** |

1 2 3 4 5 6 7

5) Dado el conjunto de datos A, escribir el conjunto de datos B

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | | Mes | X | | 1 | 15 | | 2 | 5 | | 3 | 20 | | 4 | 15 | | 5 | 10 | | 6 | 5 | | Proc rank data=A out=B condense ;  Var x; ranks rrx; | |  |  |  | | --- | --- | --- | | Mes | X | Rrx | | 1 | 15 | 3 | | 2 | 5 | 1 | | 3 | 20 | 4 | | 4 | 15 | 3 | | 5 | 10 | 2 | | 6 | 5 | 1 | |

b)proc transpose data=A out=B name=variable prefix=mes;

var x; id mes; run;

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Variable | Mes1 | Mes2 | Mes3 | Mes4 | Mes5 | Mes6 |
| X | 15 | 5 | 20 | 15 | 10 | 5 |

6) El conjunto de datos Ozono tiene información diaria (variable fecha) de cada estación (variable estacion) de la cantidad máxima (ozono\_max) y minima (ozono\_min) de ozono en el aire. Así mismo se indican las variables hora\_max y hora\_min indicando la hora (entre 1 y 24) en las que se alcanzaron esos valores extremos y la variable comunidad que contiene la Comunidad Autónoma donde se encuentra la estación. Se pide:

a) Escribir un programa que cree un conjunto de datos (denominado A) que contenga el porcentaje de veces que en una de las 24 horas se ha producido la cantidad máxima de ozono en una comunidad (variables comunidad, hora\_max y la variable que refleje el porcentaje pedido).

Proc freq data=ozono;

Tables comunidad\*hora\_max / out=A (keep=comunidad hora\_max pctrow);

Run;

b) Utilizando el conjunto creado anteriormente, escribir el codigo que represente un diagrama de tartas en donde cada sector circular sea proporcional a las veces que se alcanza el maximo de ozono en una hora para cada comunidad, intentando representar en cada página 6 comunidades diferentes y agrupando aquellas horas que recojan menos del 4% de los casos.

Proc gchart data=A;

Pie horas\_max / discrete group=comunidad others=4 down=2 across=3 sumvar=pctrow;

Run;

c) Escribir las sentencias que creas necesarias para construir un diagrama de bigotes y cajas para la variable que contiene el valor maximo del ozono por comunidad.

Proc sort data=ozono; by comunidad; run;

Proc boxplot data=ozono;

Plot ozono\_max\*comunidad;

Run;

\*

**Solucion 3º control A2**

1) ¿Qué acciones se deducen a partir del siguiente código?:

Proc copy in=a out=b move; run;

Que todos los conjuntos de datos y catálogos que se encuentran en la librería a se desplazan a la librería b, quedando la librería a vacia.

2) Dado el conjunto de datos A y el conjunto de datos B, crear el conjunto de datos C

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  | | --- | --- | --- | | **A** | | | | **Comunidad** | **Grupo** | **Destino** | | Andalucia | Zeta | New York | | Andalucia | Beta | Chicago | | Baleares | Riva | Atenas | | Baleares | Zote | Lisboa | | |  |  |  | | --- | --- | --- | | **B** | | | | **Comunidad** | **Origen** | **Precio** | | Andalucia | Jerez | 1000 | | Baleares | Manacor | 5000 | | Galicia | Tuy | 12000 | |

a) Data C; merge A(in=a) B (in=b) ; by Comunidad; z=a+b; if z>1; run;

| **Comunidad** | **Grupo** | **Destino** | **Origen** | **Precio** | **z** |
| --- | --- | --- | --- | --- | --- |
| Andalucia | Zeta | New York | Jerez | 1000 | 2 |
| Andalucia | Beta | Chicago | Jerez | 1000 | 2 |
| Baleares | Riva | Atenas | Manacor | 5000 | 2 |
| Baleares | Zote | Lisboa | Manacor | 5000 | 2 |

b) Data C; set B; suma+precio; drop comunidad origen; dato=comunidad || origen; run;

| **Precio** | **suma** | **dato** |
| --- | --- | --- |
| 1000 | 1000 | AndaluciaJerez |
| 5000 | 6000 | Baleares Manacor |
| 12000 | 18000 | Galicia Tuy |

3) Indicar el contenido del conjunto de datos A;

Data A; input x y @ ; do i=1 to y; input z @; w=(z>y) ; output; run;

3 2 1 7 3

4 1 2 1

| **x** | **y** | **i** | **z** | **w** |
| --- | --- | --- | --- | --- |
| 3 | 2 | 1 | 1 | 0 |
| 3 | 2 | 2 | 7 | 1 |
| 4 | 1 | 1 | 2 | 1 |

4) Indica el lugar en el que pueden aparecer las siguientes sentencias u opciones, y si existe algún procedimiento exclusivo al que pertenezcan. Si es una opción identificar la sentencia a la que pudiera pertenecer.

a) Symbol i=needle v=dot h=3 w=4;

Sentencia que puede ir en cualquier lugar no tiene que pertenecer a ningún procedimiento. Se utiliza en las salidas graficas donde se representan unas funciones frente a otras (sgplot, gplot, plot,…).

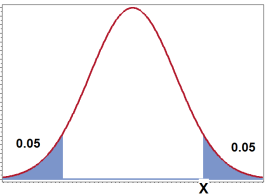
b) ppplot z; sentencia del procedimiento Univariate

c) cdfplot x; sentencia del procedimiento Univariate

d) spearman; opcion del procedimiento proc corr. Pertenece a la sentencia Proc corr

e) table (a b) (c all), x\*mean; sentencia del procedimiento Tabulate

5) Escribir la sentencia que permite determinar el valor x tal que verifique que la probabilidad de que una ley normal de media 0 y desviacion tipica 2 sea mayor que |x| sea igual a 0.1.

X=quantile(‘normal’,0.95,0,2); 

6) El conjunto de datos Ozono tiene información diaria (variable fecha) de cada punto de muestreo (variable punto) de la cantidad máxima (ozono\_max) y minima (ozono\_min) de ozono en el aire. Asi mismo se indican las variables hora\_max y hora\_min indicando la hora (entre 1 y 24) en las que se alcanzaron esos valores extremos y la variable comunidad que contiene la Comunidad Autonoma donde se encuentra la estación. Se pide:

a) Escribir el código que permita crear un diagrama de barras vertical para cada estación del año, donde cada barra represente a una comunidad y su altura proporcional al valor medio de la variable ozono\_min. (las estaciones del año son enero-marzo invierno, abril-junio primavera, julio-septiembre en verano y octubre-diciembre otoño). Un valor posible de la variable es 17MAR2018 (variable numerica).

Data nueva; set ozono; mes=month (fecha);

if mes in (1,2,3) then season=’invierno’;

if mes in (4,5,6) then season=’primavera’;

if mes in (7,8,9) then season=’verano’;

if mes in (10,11,12) then season=’otoño’;

run;

proc gchart data=nueva;

vbar comunidad / sumvar=ozono\_min type=mean group=season;

run;

b) Escribir el código que permita representar gráficamente la evolución a lo largo del tiempo (las fechas) de los maximos diarios y minimos diarios para cada punto de muestreo (una figura para cada punto de muestreo). Esas dos cantidades a representar tienen que tener escalas diferentes en los gráficos.

Proc gplot plot=ozono;

By punto;

Plot ozono\_max\* fecha / legend;

Plot2 ozono\_min\*fecha /legend;

Run;