**SOFTWARE I CONVOCATORIA DE JUNIO**

1)Escribir las sentencias que creas necesarias para crear un conjunto de datos permanente llamado C, a partir del fichero valores.txt guardado en la carpeta C:\misdocumentos (se compone de las líneas que tenemos a continuación, donde cada línea es una observación diferente): ). Pon el nombre que desees a las 4 variables.

Washington 178 M 3.5

Leverkusen 111 M 2.4

Burdeos 271 M 45

Libname e ‘e:’;

Data e.c;

 Infile ‘c:\misdocumentos\valores.txt’;

 Format ciudad $10. ;

Input ciudad $ x1 a $ x2;

Run;

2) Se tiene el fichero excell asignatuas.xml guardado en la unidad de disco e:\ . Crear el conjunto de datos SAS asignaturas y guardarlo en C:\Mis documentos.

Libname c ‘c:\mis documentos’;

Proc import datafile=’e:\asignaturas.xml’ out=c.asignaturas dbms=xml; run;

3) identifica el conjunto de datos pp

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Data pp;  Input x y @@;  Do while (x>y);  X=x-1;  Z= y-x;  Output;  End; Datalines;5 7 6 3  |

| **Obs** | **x** | **y** | **Z** |
| --- | --- | --- | --- |
| **1** | 5 | 3 | -2 |
| **2** | 4 | 3 | -1 |
| **3** | 3 | 3 | 0 |

 |

4) Indentificar el contenido del conjunto de datos pp;

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Data pp;**  **Input z1 z2 z3 z4 @@;**  **Drop z2;**  **If z2>z3 then output;**  **Cards;**  **12 23 55 18 21 33 47 25 12 37 42 11**  **15 34 21 17 11 12 28 11 26 43 41 17**  |

| **Obs** | **z1** | **z3** | **z4** |
| --- | --- | --- | --- |
| **1** | 15 | 21 | 17 |
| **2** | 26 | 41 | 17 |

 |

5) Dado los conjuntos de datos A y B. Crear el conjunto de datos C

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|

|  |
| --- |
| A |
| Comunidad | Pib\_percapita | colegios |
| Andalucia  | 18000 | 4500 |
| Cataluña | 30000 | 4000 |
| Extremadura | 15000 | 500 |

 |

|  |
| --- |
| B |
| Comunidad | Profesorado |
| Andalucia | 4400 |
| Extremadura | 5000 |
| Pais Vasco | 6000 |

 |

1. Data C ; merge A B; by Comunidad, run;

| **Obs** | **comunidad** | **Pib\_percapita** | **colegios** | **Profesorado** |
| --- | --- | --- | --- | --- |
| **1** | Andalucia | 18000 | 4500 | 4400 |
| **2** | Cataluña | 30000 | 4000 | . |
| **3** | Extremadura | 15000 | 500 | 5000 |
| **4** | Pais Vasco | . | . | 6000 |

b) Data C; merge A B; drop colegios; run;

| **Obs** | **comunidad** | **Pib\_percapita** | **Profesorado** |
| --- | --- | --- | --- |
| **1** | Andalucia | 18000 | 4400 |
| **2** | Extremadura | 30000 | 5000 |
| **3** | Pais Vasco | 15000 | 6000 |

c) data c; set b; profesorado=profesorado/1000; where profesorado<6000; run;

| **Obs** | **comunidad** | **Profesorado** |
| --- | --- | --- |
| **1** | Andalucia | 4.4 |
| **2** | Extremadura | 5.0 |

d) Data c; set a; razon= Int( pib\_percapita/colegios); if razon>10 then output; run;

| **Obs** | **comunidad** | **Pib\_percapita** | **colegios** | **Razón** |
| --- | --- | --- | --- | --- |
| **1** | Extremadura | 15000 | 500 | 30 |

6) Indicar el contenido del conjunto de datos A:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Data A;  Input x y @;  Do z=1 to 3;  Input w;  Output;  End; Cards; 3 5 6 12 34 22 33 44 32 3 4 5 6 7 1 |

| **Obs** | **x** | **y** | **z** | **w** |
| --- | --- | --- | --- | --- |
| **1** | 3 | 5 | 1 | 6 |
| **2** | 3 | 5 | 2 | 2 |
| **3** | 3 | 5 | 3 | 1 |

 |
| 7) A través de un único paso Data. Crear un programa que permita calcular la suma y la media de los costes que se han producido a lo largo de un periodo de tiempo partiendo de la información del conjunto de datos SAS Tienda (presente en la librería work).(Nota: este nuevo conjunto sólo tiene que tener una única observación) |

|  |
| --- |
| Tienda |
| Mes | Coste |
| Enero | 1011 |
| Febrero | 1490 |
| Marzo | 989 |
| Abril | 1290 |
| Mayo | 1096 |
| Junio | 987 |
| Julio | 1789 |
| Agosto | 2015 |

  |

**data** tiendab;

 set tienda end=pp;

 retain sumcoste (**0**);

 sumcoste=sumcoste+coste;

 n+**1**;

 if pp=**1** then do;

 media=sumcoste/n;

 output;

 end;

**run**;

8) Se realiza una encuesta entre hortelanos para estimar el número de kilos cosechados de tomates, calabacines, guisantes y zanahorias. Si un hortelano no cultiva alguno de los productos el valor que se le asigna es 0. La información está contenida en el conjunto de datos SAS denominado Huertas que tiene el siguiente formato:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Nombre | Tomates | Calabacines | Guisantes | Zanahorias |
| Antonio | 10 | 2 | 40 | 0 |
| Luis | 15 | 5 | 10 | 80 |
| …. | …. | …. | …. | …. |

Crear el conjunto de datos resumen que contenga sólo la siguiente información: el nombre del hortelano, el número total de kilos cosechados por cada hortelano, el porcentaje sobre el total del peso de los tomates y el numero de productos cultivados.

Data resumen;

 Set huertas;

 Total=tomates+calabacines+guisantes+zanahorias;

 Perc\_tomates=tomates/total;

 N\_productos=(tomates>0)+ (calabacines>0) + (guisantes>0)+ (zanahoras>0);

 Keep nombre total perc\_tomates n\_productos;

Run;

9) Escribir un programa o sentencia que permita calcular el valor x que verifica que la probabilidad de que una ley Normal (media=2, desviación típica=3) sea mayor que x vale 0.2.

X=quantile(‘Normal’,0.8,2,3);

10) Calcular la probabilidad de que una variable aleatoria que se distribuye según una ley binomial de parametros n=20 y p=0.2 sea mayor que 5 y menor que 12.

Prob=cdf(‘binomial’,11,0.2,20)-cdf(‘binomial’,5,0.2,20);

11) Los valores de la longitud en centímetros de 20 cables eléctricos se encuentran en el fichero de texto cables.txt dentro del subdiretorio c: Este fichero tiene la siguiente forma:

100 45 54 67 34 43 22 46 58 89

89 12 109 105 34 76 34 88 99 12

a) Crear el conjunto de datos SAS cables conteniendo a la variable longitud leída en metros.

Data cables;

Infile ‘c:\cables.txt’;

Input longitud @@;

Longitud=longitud/100;

run;

b) Calcular sus estadísticos más habituales (media, desviación típica, máximo y mínimo).

c) Realizar una gráfica ppplot que permita contrastar gráficamente si los cables se distribuyen según una ley normal.

Apartados b y c la vez;

 Proc univariate data=cables;

 Var longitud;

 Ppplot cables / normal;

Run;

12)

a) Escribir un programa que permita calcular la función de densidad de un chi-cuadrado con g grados de libertad (donde g varía desde 5 hasta 50 de 5 en 5 g.libertad) para valores x que vayan desde 0.5 hasta 100 de 0.5 en 0.5. El conjunto de datos tiene que tener por tanto esta forma: (fx se corresponde con la función de densidad de x para un chi-cuadrado con g grados de libertad)

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|

|  |  |  |
| --- | --- | --- |
| G | X | Fx |
| 5 | 0.5 | 0.03662 |
| 5 | 1 | 008066 |
| … | … | … |
| 5 | 100 | 0 |
| 10 | 0.5 | 0.00006 |
| …. | …. | … |
| 50 | 100 | 0.00001 |

 | Data función;  Do g=5 to 50 by 5; Do x=0.5 to 100 by 0.5;  fx=pdf(‘chisquare’,c,g);  output;  end; end; |

b) Escribir el código que dibuje todas esas funciones chi-cuadrado en la misma figura (una curva para cada valor de g).

proc gplot;

 plot fx\*x=g;

run;

13) Escribir el conjunto de datos ramplón que se obtiene de la siguiente secuencia de código.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Data** ramplon;  Input x y;  If x>y then x=y+**1**;  Else output; cards; 3 5 2 7 4 1  |

| **Obs** | **x** | **y** |
| --- | --- | --- |
| **1** | 3 | 5 |
| **2** | 2 | 7 |

 |

14) identificar al conjunto de datos pp;

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Data pp; retain x(6);**  **Do i=1 to 6;**  **Y+1;**  **X=x+y;**  **Output;**  **End;**  |

| **Obs** | **x** | **i** | **Y** |
| --- | --- | --- | --- |
| **1** | 7 | 1 | 1 |
| **2** | 9 | 2 | 2 |
| **3** | 12 | 3 | 3 |
| **4** | 16 | 4 | 4 |
| **5** | 21 | 5 | 5 |
| **6** | 27 | 6 | 6 |

 |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 15) Sabiendo que las líneas de texto que siguen a la sentencia cards en cada uno de los siguientes apartados es siempre la misma y concretamente: 34 21 21 10 8 6 5 21 2 12 3 4 3 2 4 5 31 3 Se pide escribir el contenido del archivo *persa.sas7bdat* almacenado en la librería *work* en los siguientes casos: 1. Data persa; do j=1 to 5;

 input casas @ ;  output;  end;  output;  cards; |

| **Obs** | **j** | **casas** |
| --- | --- | --- |
| **1** | 1 | 34 |
| **2** | 2 | 21 |
| **3** | 3 | 21 |
| **4** | 4 | 10 |
| **5** | 5 | 8 |
| **6** | 6 | 8 |
| **7** | 1 | 12 |
| **8** | 2 | 3 |
| **9** | 3 | 4 |
| **10** | 4 | 3 |
| **11** | 5 | 2 |
| **12** | 6 | 2 |

 |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1. Data persa ; do j=1 to 5; input casas @@; end; output;

 input pepe; output; cards; |

| **Obs** | **j** | **casas** | **pepe** |
| --- | --- | --- | --- |
| **1** | 6 | 8 | . |
| **2** | 6 | 8 | 6 |
| **3** | 6 | 2 | . |
| **4** | 6 | 2 | 4 |

 |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1. Data persa ;

 do j=1 to 5;  input casas @;  output;  end; cards; |

| **Obs** | **j** | **casas** |
| --- | --- | --- |
| **1** | 1 | 34 |
| **2** | 2 | 21 |
| **3** | 3 | 21 |
| **4** | 4 | 10 |
| **5** | 5 | 8 |
| **6** | 1 | 12 |
| **7** | 2 | 3 |
| **8** | 3 | 4 |
| **9** | 4 | 3 |
| **10** | 5 | 2 |

 |

1. Data persa; do j=1 to 5; input casas; end; cards;

 No hay conjunto de datos a leer

16) Tenemos el siguiente conjunto de datos A, con las notas de diferentes controles realizados a los alumnos

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Alumno | Control1 | Control2 | Control3 | Control4 | Control5 | Control6 |
| Juan | 5 | 3 | 6 | 2 | 7  | 9 |
| Pedro | 9 | 8 | 7 | 8 | 9 | 6 |
| …. | …. | …. | …. | …. | …. | … |

Suponer que el profesor decide incrementar en un 10% las notas de todos los controles de todos los alumnos. Crear el código que pueda modificar las notas. Usar sentencia array que facilite la modificación de las variables.

Data a;

Set a;

Array control(6);

Do i=1 to 6;

 Control(i)= control()\*1.1;

End;

Run;

17) Identificar al conjunto de datos A;

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Data A;  Input x @@; Retain y (0); If y>x then x=x+1; Else y=y+x; Cards; 3 7 2 5  |

| **Obs** | **x** | **y** |
| --- | --- | --- |
| **1** | 3 | 3 |
| **2** | 7 | 10 |
| **3** | 3 | 10 |
| **4** | 6 | 10 |

 |

18) En el conjunto de datos país, se tienen una observación por año conteniendo las variables year (el año desde 1978 hasta 2017), parados (con el total de parados en cada año), deuda (con la deuda publica en millones de euros) y presos (con el número de personas que se encuentran privadas de libertad en la carcel).

a) Crear un programa que permita representar una figura con la evolución de las 3 variables (parados, deuda, presos) a lo largo de los años. Las 3 curvas en la misma figura.

 Proc gplot data=país;

 Plot (parados deuda presos)\*year / OVERLAY;

Run;

1. Incluir una variable que se llame ideología, de tal forma que valga derecha en los años que gobernó la derecha en el país (1978-1982, 1996-2004, 2012-1017) e izquierda cuando gobernó el PSOE (resto de años). Crear histogramas horizontales que representen la distribución de los parados y de presos para cada valor de la variable ideología.

Data pais2; set país;

 If (país >1983 and país<1996) or (país>2004 and pais<2012) then ideologia=’izquierda’;

 Else ideologia=’derecha’

Run;

Proc gchart;

 Hbar parados presos / group=ideologia;

Habar

c) Escribir el programa que permita calcular la media anual de parados, presos y deuda según el tipo de partido que gobernó en España. Crear un conjunto de datos nuevo con esa información (una observación por valor de ideología).

Proc means data=país2 nway;

Class ideología;

Var parados presos deudas;

Output out=pais3 mean= ;

Run;

19) El fichero de datos SAS alumnos (en librería work) contiene el sexo (‘M’ o ‘F’) la edad (en meses) el peso (en kg) y la talla (en cm,) de los alumnos de una Facultad. Se pide:

a) Crear un conjunto de datos que contenga la razón entre la talla y el peso, además de la edad en años (y sin decimales) que tienen todos ellos.

Data a; set alumnos;

Razón=talla/peso;

Edad=int(edad/12);

b) A partir del conjunto de datos anterior crear otro conjunto de datos en el que se seleccione solo a los varones con una razón mayor que dos.

Data b; set a;

Where razón>2 and sexo=’M’;

Run;

c) Ordenar los datos del conjunto creado en el apartado a) por sexo y por talla.

Proc sort data=b;

By sexo talla;

Run;

d) Crear un conjunto de datos que contenga la media del peso y la talla para cada sexo.

Proc means data=b;

By sexo;

Var peso talla;

Output out=c mean= autoname;

Run;

e) Crear un codigo SAS que permita visualizar los histogramas de la talla y el peso para cada sexo de forma diferenciada. Contrastar que siguen una ley de distribución Normal cada una de esas variables para cada sexo.

Proc univariate data=b;

Var talla peso;

Histograma talla peso /normal;

Class sexo;

Run;

f) Escribir el código de un procedimiento SAS que permita visualizar la media, la desviación típica, el coeficiente de kurtosis, el coeficiente de asimetria para el peso de todos los alumnos. Además se pretende obtener un intervalo de confianza para la media del 95% sobre esa variable.

Proc univariate data=b cibasic ; var peso; run; o bien:

Proc means data=b mean std kurtosis skewness clm;run;

20) En el conjunto de datos Empresa, tenemos los salarios, impuestos (IRPF) y cotizacion (de la Seguridad Social) de cada empleado desde 1998 hasta 2017 (variable Year). Se pide:

a) Crear una nueva variable denominada Coste\_individual que sea la suma de los salarios, impuestos y cotización para cada empleado.

Data empresa; set empresa;

 Coste\_individual=salarios + impuestos + cotización; run;

b) Crear el conjunto de datos SAS que contenga los totales anuales de las variables: salarios, impuestos y cotizacion, con el volumen total pagado por la empresa en esas 3 partidas y el porcentaje que le corresponde cada año a cada partida.

Proc means data=empresa;

Var salarios impuestos cotización coste\_individual;

By year;

Output out=suma sum=;

Run;

Data suma; set suma; drop \_freq\_ \_type\_;

 Pct\_salarios=salarios/ coste\_individual;

 Pct\_impuestos=impuestos/ coste\_individual;

 Pct\_cotizacion= cotización/ coste\_individual;

Run;

d) A partir del conjunto creado en b y usando un bucle calcular los valores estimados en los años posteriores a 2017 (y hasta 2030) de los montantes totales de esas 3 variables económicas suponiendo incrementos anuales del 1%, 1.5% y 2% respectivamente.

Data simula;

Set suma (end=xx);

If xx=1 then do;

 Do year=2018 to 2030;

 salarios=salarios\*(1.01);

 impuestos=impuestos\*(1.015);

 cotizacion=cotizacion\*(1.02);

 output;

 end;

end;