

Ejercicios de CREACION DE VARIABLES Y PROCS

Los archivos necesarios para hacer esta actividad se encuentran en la carpeta "dades_s09" que se creará automáticamente al descomprimir el archivo "sesion9.zip". Se recomienda crear en la unidad de trabajo (D:\, F:\, etc...) la estructura de carpetas siguiente:

D:\CURSAS16_17\S09\dades_s09\ que contendrá las bases de datos de la sesión de trabajo, y

D:\CURSAS16_17\S09\Ejercicios\ que contendrá las bases de datos que se generen al resolver esta actividad.

Ejercicio 1. CREACIÓN DE VARIABLES

1. Crea una base temporal en D:\CURSAS16_17\S09\dades_s09\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.

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.

Y9: si X1 es positiva, Y9 será la raíz cuadrada de X1.

si $X1 \leq 0$ y $X3 < 0.5$ Y9 será la raíz cuadrada de X3.

si $X1 > 0$ y $X3 \geq 0.5$ Y9 será $\cos(X3)$.

Guardar la sintaxis como Ej1_S09.SAS.

Ejercicio 2. CREACIÓN DE VARIABLES Y PROCEDIMIENTOS

El archivo D:\CURSAS16_17\S09\dades_s09\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?
4. Crea la base de datos D:\CURSAS16_17\S09\dades_s09\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.

a) ¿Qué porcentaje de individuos incongruentes hay?

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

c) ¿Qué porcentaje de los que gastan más de un 25% en vivienda gastan menos de un 15% en ocio?

Guardar la sintaxis como Ej2_S09.SAS.

Ejercicio 3. CREACIÓN DE VARIABLES Y PROCEDIMIENTOS

La base de datos D:\CURSAS16_17\S09\dades_s09\SOCIS.SAS7BDAT contiene información relativa a los socios de una entidad deportiva. A partir de esta base, crear la base de datos D:\CURSAS16_17\S09\dades_s09\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.
2. ¿Cuál es la cuota media de las mujeres? ¿y de los hombres?
3. ¿Qué antigüedad tiene la mujer que más años lleva como socia? ¿y el hombre que más años lleva como socio?
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_S09.SAS.

Ejercicio 4. CREACIÓN DE VARIABLES Y PROCEDIMIENTOS

El archivo D:\CURSAS16_17\S09\dades_s09\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:\CURSAS16_17\S09\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_S09.SAS.

Ejercicio 5. CREACIÓN DE VARIABLES Y PROCEDIMIENTOS

El archivo D:\CURSAS16_17\S09\dades_s09\SINIESTROS.SAS7BDAT contiene las siguientes variables:

<u>Variable</u>	<u>Descripcion</u>
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?
3. ¿Qué porcentaje de mujeres no ha declarado ningún siniestro? ¿y de hombres?
4. Haz un histograma de la variable COSTE con curva normal superpuesta y calcula los cuartiles.
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:\CURSAS16_17\S09\dades_s09\SINIESTROS2.SAS7BDAT
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_S09.SAS.

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:

$Y = \sqrt{\text{valor absoluto de } X}$ si el valor de X es negativo

$Y = \sqrt{\text{valor de } X}$ si el valor de X es positivo

Guardar la sintaxis como Ej6_S09.SAS.

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.
- 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 missins superior a 53?