

# Computación Estadística

## EPG3308

Profesora: María Inés Godoy,  
Ayudante: María De Los Angeles Villena

Primer Semestre 2017:21 marzo

# Opciones de DATA

Comandos comunes dentro de una DATA son;

- ▶ **DROP:** Elimina variables
- ▶ **KEEP:** Guarda en la DATA sólo las variables que se desean mantener en ellas.
- ▶ **RENAME:** (Nombre-Antiguo = Nombre-Nuevo) Renombra las variables
- ▶ **FIRSTOBS:** (FIRSTOBS=n) Indica desde que línea se comienza a leer.
- ▶ **OBS:** (OBS=n) Indica en que línea termina la lectura de datos.
- ▶ **WHERE:** (WHERE=Condición) Filtra por filas y mantiene sólo las que cumple la condición.

# DATA 1

Nacidos vivos inscritos según peso al nacer, por región y comuna.

Chile, 2011

```
DATA PESO;  
infile "C:\Peso2011.txt"  
delimiter=';' firstobs=2;  
Input Region $ : 39.  Comuna $ : 20.  
Total MENOS1500  
ENTRE_1500A2499  
ENTRE_2500A2999  
Mayor_igual_3000  
Ignorado;  
Run;
```

## DATA 2

Nacidos vivos inscritos según edad de la madre, por región y comuna.  
Chile, 2011

```
DATA EDAD;  
infile "C:\Edad2011.txt"  
dlim=';' firstobs=2;  
Input Region $ : 39.  
Comuna $ : 20.  
Total Menor_15  
Entre15_19  
Entre20_34  
Mas_Igual35  
Ignorado;  
Run;
```

# Ejemplos

```
DATA PESO2 (DROP=Total Ignorado);  
SET PESO;  
PROC PRINT;  
RUN;
```

```
DATA EDAD2 (KEEP =rREGION COMUNA MENOR_15 ENTRE20_34 MAS_iGUAL35);  
SET EDAD;  
PROC PRINT;  
RUN;
```

# Ejemplos

```
DATA PESO_EJEMPLO;  
SET PESO2 (FIRSTOBS=2 OBS=5);  
PROC PRINT;  
RUN;
```

```
DATA PESO_REGION_M;  
SET PESO2 (Where=(Region='Metropolitana de Santiago'));  
PROC PRINT;  
RUN;
```

```
DATA PESO_REGION_3;  
SET PESO2 (Where=(Region='De Atacama'));  
PROC PRINT;  
RUN;
```

# Ejemplos

```
DATA PESO_REGION_M_Sant;  
SET PESO2 (Where=(Region='Metropolitana de Santiago' and  
Comuna='Santiago'));  
PROC PRINT;  
RUN;
```

# Concatenar dos data.sas

Se usa la opción **SET**. La Base de datos tendrá tantas observaciones como la suma de las observaciones de los archivos a concatenar.

```
DATA PESO_REGION_3_y_M;  
SET PESO_REGION_M PESO_REGION_3;  
PROC PRINT;  
RUN;
```



## Unir dos archivos

Se usa la opción **MERGE**. La Base de se une por la variable tipo, la cual debe ser común en ambos y ordena previamente.

```
PROC SORT DATA=PESO; BY Region Comuna;RUN;
```

```
PROC SORT DATA=EDAD; BY Region Comuna;RUN;
```

```
DATA UNION;
```

```
MERGE Edad Peso;
```

```
By Region Comuna;
```

```
PROC PRINT;
```

```
RUN;
```

# Unir dos archivos

```
PROC SORT DATA=PESO_REGION_M; BY Region Comuna;RUN;  
data UNION2;  
merge PESO_REGION_M(IN=k) Edad(IN=j);  
by Region Comuna;  
if k=j;  
PROC PRINT;  
RUN;
```

# Modificación y creación de variables.

Para crear variable, sólo se agrega una linea dentro del data, tal que

**Nueva Variable= Expresión**

Agunos operadores aritméticos son;

- ▶ Potenciación: \*\*
- ▶ Multiplicación:\*
- ▶ División:
- ▶ Suma: +
- ▶ Resta: –

```
DATA Ejemplo;
```

```
Set Union2;
```

```
Total_nac_ci=MENOS1500 + ENTRE_1500A2499 + ENTRE_25
```

```
Total_mad_ci = Menor_15 + Entre15_19 + Entre20_34 +
```

```
RE1 = Menor_15/Total_mad_ci*100;
```

```
RE2 = Entre15_19/Total_mad_ci*100;
```

```
RE3 = Entre20_34/Total_mad_ci*100;
```

```
RE4 = Mas_Igual35/Total_mad_ci*100;
```

```
proc print;
```

```
run;
```

# Sentencias de Control

- ▶ **IF**; (IF condición).
  - ▶ Igual a: EQ o =
  - ▶ Diferente a: NE
  - ▶ Menor que: LT o <=
  - ▶ Mayor que: GT o >=
  - ▶ Algunos operadores lógicos: OR, AND, NOT

## Ejemplo

A partir del archivo Ejemplo, que ya creamos, Crearemos una DATA que contenga sólo a las comunas de la región Metropolitana que al menos el 50 % de las madres tenga entre 20-34 años durante el nacimiento de sus hijos en el año 2011.

```
DATA Ejemplo2;  
SET Ejemplo (Where=(Region='Metropolitana de Santiago'));  
if RE3 >= 50;  
proc print;  
run;
```

```
DATA Ejemplo2;  
SET Ejemplo;  
if RE3 >= 50 and Region='Metropolitana de Santiago';  
proc print;  
run;
```

## Ejercicio en clases

Crear una data, donde estén todas las comunas a lo largo de Chile, con las categorías de peso del niño al nacer y de la edad de la madre.

Guarde esta en una librería creada por usted.

- ▶ Crear una nueva variable que indique la categoría de peso al nacer más frecuente,
- ▶ Crear una nueva variable que indique la segunda categoría de peso al nacer más frecuente,
- ▶ Crear una nueva variable que indique la tercera categoría de peso al nacer más frecuente,
- ▶ Crear una nueva variable que indique la categoría de peso al nacer menos frecuente,

## Ejercicio en clases

- ▶ Repita los pasos anteriores con las categorías de edad de la madre.
- ▶ Realice una descripción de los nacimientos en Chile durante el 2011 (AÚN NO TENEMOS LAS HERRAMIENTAS PARA RESOLVER).
  - ▶ Que región presenta más nacimiento de madres adolescentes en el año 2011 en Chile.
  - ▶ Qué región presentó durante el 2011 nacimientos con bebés de bajo peso (menos 1500gr) con mayor frecuencia a lo largo de Chile. (considere el total de nacimientos en la región)