

Computación Estadística

EPG3308

Profesora: María Inés Godoy,

Ayudante: María De Los Angeles Villena

Primer Semestre 2017: 4 abril

PROC SUMMARY

Es una herramienta útil para calcular los estadísticos descriptivos de un conjunto de datos. Este procedimiento es muy similar al PROC MEANS.

Las sintaxis de este procedimiento, corresponde a:

PROC SUMMARY <options> ;

BY <DESCENDING> variables ;

CLASS variables </options>

FREQ variable ;

ID variable(s) ;

PROC SUMMARY

OUTPUT <OUT=SAS-data-set> <output-statistic-specification(s)>
<id-group-specification(s)> <maximum-id-specification(s)>
<minimum-id-specification(s)> </ option(s)>;

TYPES request ;

VAR variables </ WEIGHT=weight-variable>;

WAY list;

WEIGHT variable ;

PROCEDIMIENTO SUMMARY

- ▶ **PROC SUMMARY** Invoca el procedimiento, calcula estadísticos descriptivos a través de todas las variables o un grupo de observaciones.
- ▶ **BY** Usted lo especifica si desea calcular los estadísticos separados por grupo.
- ▶ **CLASS** Para especificar variables cuyos valores definen sub grupos para el análisis.
- ▶ **FREQ** Para especificar la variable cuyos valores especifica la frecuencia para cada observación.
- ▶ **ID** Incluye variables de identificación adicionales en el conjunto de salida.

PROCEDIMIENTO SUMMARY

- ▶ **OUTPUT** Crea un conjunto de datos de salida que contiene las estadísticas especificadas y variable de identificación.
- ▶ **TYPES** Identifica las combinaciones específicas de las variables de clase que se usan para subdividir los datos.
- ▶ **VAR** Especifica la variables de análisis y su orden en los resultados.
- ▶ **WAYS** Especifica el número de maneras de hacer combinaciones únicas de las variables de clase.
- ▶ **WEIGHT** Identifica una variable cuyos valores son los peso de cada observación en los cálculos estadísticos.

PROCEDIMIENTO SUMMARY

```
proc summary data = sashelp.class(where=(age in (12,14,15)));  
class sex age;  
var height;  
output out=ejemplo1 n=hgt_n mean=hgt_mean std=hgt_std;  
run;
```

PROCEDIMIENTO SUMMARY

```
proc means data = sashelp.class(where=(age in (12,14,15)));  
class sex age;  
var height;  
output out=ejemplo2 n=hgt_n mean=hgt_mean std=hgt_std;  
run;
```

PROCEDIMIENTO SUMMARY

```
proc summary data = sashelp.class;  
class sex age;  
var height;  
output out=ejemplo3  
mean=hgt_mean  
max=hgt_max  
min=hgt_min  
maxid(height (name))=maxhtname  
minid(height (name))=minhtname;  
run;
```


PROC TABULATE

El procedimiento TABULATE, muestra las estadísticas descriptivas en forma tabular, usando una o más variables del conjunto de datos. Calcula muchas de los estadísticos que calculan PROC MEANS, PROC FREQ. LA diferencia es que este procedimiento ofrece métodos simples para crear informes tabulares. La sintaxis corresponde a;

PROCEDIMIENTO TABULATE

```
PROC TABULATE <options> ;  
BY <DESCENDING> variable-1  
  < ...<DESCENDING> variable-n  
  <NOTSORTED> ;  
CLASS variable(s) </ options >;  
CLASSLEV variable(s) / STYLE=<style-element-name | PARENT>  
  < style-attribute-specification(s) >; FREQ variable ;  
KEYLABEL keyword-1='description-1'  
  <...keyword-n='description-n'>;
```

PROCEDIMIENTO TABULATE

KEYWORD keyword(s) / **STYLE**=<style-element-name | PARENT>

< style-attribute-specification(s) >;

TABLE <<page-expression,> row-expression,> column-expression< /
table-option(s)>; ;

VAR variables </ options >;

WEIGHT variable ;

PROCEDIMIENTO TABULATE

- ▶ **PROC TABULATE** Invoca el procedimiento, calcula estadísticos descriptivos y los muestra en formato de tabla.
- ▶ **BY** Usted lo especifica si desea calcular los estadísticos separados por grupo.
- ▶ **CLASS** Para especificar variables cuyos valores definen sub grupos para el análisis.
- ▶ **CLASSLEV** Especifique un estilo para los niveles de las variables de clase.
- ▶ **FREQ** Identificar una variable en el conjunto de datos de entrada cuyos valores representan la frecuencia de cada observación.

PROCEDIMIENTO TABULATE

- ▶ **KEYLABEL** Especifica una etiqueta para **KEYWORD**.
- ▶ **KEYWORD** Especifica un estilo para el encabezado de **KEYWORD**.
- ▶ **TABLE** Describe la tabla creada.
- ▶ **VAR** Especifica la variables de análisis.
- ▶ **WEIGHT** Identifica una variable cuyos valores son los peso de cada observación en los cálculos estadísticos.

PROCEDIMIENTO TABULATE

EL siguiente ejemplo corresponde a PROC TABULATE. Corresponde a un conjunto de datos que contiene los gastos de energía de clientes residenciales y de empresa de diferentes regiones y sectores del Norte de Chile.

```
data Energia;  
length Estado $2;  
input Region Provincia Estado $ Tipos Gastos;  
datalines;  
1 1 ME 1 708  
1 1 ME 2 379  
1 1 NH 1 597  
.  
4 4 HI 2 298  
;
```

PROCEDIMIENTO TABULATE

```
proc tabulate data=energia format=dollar12.;  
class region provincia estado Tipos;  
var Gastos;  
table Region*Provincia, Tipos*Gastos /printmiss;;  
run;
```

```
*OPCIONAL;  
proc format;  
value fm_region 1='Tarapaca'  
2='Antofagasta'  
3='Atacama'  
4='Coquimbo';  
value fm_provincia 1='Cordillera'  
2='Pre-Cordillea'  
3='Valle'  
4='Costa';  
value fm_tipo 1='Residencial'  
2='Empresa';  
run;
```

PROCEDIMIENTO TABULATE

```
proc tabulate data=energia format=dollar12. ;  
class region provincia estado Tipos;  
var Gastos;  
table Region*Provincia, Tipos*Gastos /printmiss;  
format region fm_region. provincia fm_provincia. Tipos fm_tipo.;  
  
title 'Gastos de Energía';  
title2 '(en Pesos)';  
run;
```


PROCEDIMIENTO TABULATE

```
*SUMA por FILAS-COLUMNAS;  
ods pdf file='external-PDF-file';  
  
proc tabulate data=energia format=dollar12. ;  
class region provincia estado Tipos ;  
var Gastos;  
table Region*(Provincia all='Subtotal')  
all='Total por Regions'*f=dollar12.,  
Tipos='Clientes'*Gastos=' '*sum=' '  
all='Total clientes'*gastos=' '*sum=' ' /  
style=[bordercolor=blue];  
  
format region fm_region. provincia fm_provincia. Tipos fm_tipo.;  
title 'Gastos de Energía';  
title2 '(en Pesos)';  
run;
```

PROCEDIMIENTO TABULATE

```
ods pdf close;

ods pdf file='external-PDF-file';
proc tabulate data=energia format=dollar12. ;
class region provincia estado Tipos ;
var Gastos;
table Region*{style={background=yellow}}*(Provincia all='Subtotal')
all='Total por Region'*{style={background=red}}*f=dollar12.,
Tipos='Clientes'*Gastos=' '*sum=' '
all='Total clientes'*gastos=' '*sum=' ' /
style=[bordercolor=blue];

format region fm_region. provincia fm_provincia. Tipos fm_tipo.;
title 'Gastos de Energía';
title2 '(en Pesos)';
run; ods pdf close;
```

Ejercicio 1

El conjunto de datos Sashelp.BWeight ofrece 1.997 datos de peso al nacer del Centro Nacional de Estadísticas de Salud. El registro de datos, nacimientos simples vivos de madres entre las edades de 18 y 45 en los Estados Unidos que fueron clasificados como negro o blanco. Las variables corresponden a,

- ▶ Weight: Peso al nacer
- ▶ Black: 1 si la Mamá es negra o 0 si no.
- ▶ Married: 1 si la Mamá es casada 0 si no.
- ▶ Boy: 1 si el bebe es hombre 0 si no.
- ▶ MomAge: Edad de la madre.
- ▶ MomSmoke: 1 si la mamá fumaba 0 si no.
- ▶ CigsPerDay: Número de cigarros por día.
- ▶ MomWtGain: Peso ganado por la madre durante el embarazo.
- ▶ Visit: Visitas pre-natales
- ▶ MomEdLevel: Nivel de educación de la madre

Ejercicio 1

Con estos datos realice los siguientes ejercicios.

1. Cree una librería y guarde en ella la base de datos que se encuentra en sashelp llamada Bweight.
2. Cambie el Label de las variables, BLACK, SMOKE, WEIGHT, VISIT y MARRIED por 'COLOR DE PIEL', 'FUMADORA', 'PESO', 'Visitas al doctor' y 'Casada'; respectivamente
3. Use Proc SUMMARY para guardar el peso promedio de los recién nacidos por color de piel (black) y si la madre fumaba o no (Smoke). A simple vista que observa?
4. En el punto anterior usted debió observar que los recién nacidos de madres que fumaban en promedio pesaban menos. Realice un test de media para comprobar si es así o no. Para ello use:

Ejercicio 1

```
PROC TTEST DATA=BASE;  
class smoke;  
var weight;  
RUN;
```

- 5 Use proc tabulate para realizar la siguiente tabla.

**Peso de recién nacidos
Fuma/No Fuma**

PESO	
FUMADORA	
NO	SI
Mean	Mean
3402.31	3160.85

Ejercicio 1

6 Use proc tabulate para realizar la siguiente tabla.

Peso de recién nacidos
Color de Piel/Fuma

		PESO
		Mean
COLOR DE PIEL	FUMADORA	
Blanco	NO	3446.17
	SI	3191.45
Negro	NO	3186.77
	SI	2939.07

Ejercicio 1

- 6 Use proc tabulate para realizar la siguiente tabla.

Peso de recién nacidos
Visitas pre-natales

	PESO
	Mean
Visitas al doctor	
0	3055.33
1	3301.99
2	3276.05
3	3386.62

Ejercicio 1

6 Use proc tabulate para realizar la siguiente tabla.

Peso de recién nacidos
Visitas pre-natales/Casada

		PESO
		Mean
casada	Visitas al doctor	
NO	0	2943.77
	1	3220.31
	2	3197.83
	3	3248.20
SI	0	3241.51
	1	3386.47
	2	3374.22
	3	3431.24