Computación Estadística

EPG3308

Profesora: María Inés Godoy,

Ayudante: María De Los Angeles Villena

Primer Semestre 2017: 30 marzo

PROC UNIVARIATE

Usted puede usar PROC UNIVARIATE para obtener una variedad de estadísticos para resumir la distribución de los datos, por ejemplo,

- Momentos de la muestra.
- Medidas básicas de locación y variabilidad.
- Intervalos de confianza para la media, desviación estándar y varianza.
- Test de normalidad.
- Cuantiles, intervalos de confianza.
- Observaciones extremas, valores extremos.
- Valores missing.

PROCEDIMIENTO UNIVARIATE

PROC UNIVARIATE <options> ;

BY variables ;

CDFPLOT <variables> < / options> ;

CLASS variable-1 <(v-options)> <variable-2 <(v-options)>>

</ KEYLEVEL= value1 | (value1 value2)> ;

FREQ variable ;

HISTOGRAM <variables> < / options> ;

ID variables ;

INSET keyword-list </ options> ;

OUTPUT <OUT=SAS-data-set> <keyword1=names ...keywordk=names>

<percentile-options> ;

PPPLOT <variables> < / options> ;

PROBPLOT <variables> < / options> ;

QQPLOT <variables> < / options> ;

VAR variables ;

WEIGHT variable ;

PROCEDIMIENTO UNIVARIATE

- **PROC UNIVARIATE** Invoca el procedimiento.
- **VAR** Especifica la variable numérica para ser analizada. Si no lo especificas todas las variables numéricas de la DATA serán analizadas
- **OUTPUT** Crea un conjunto de datos de salida que contiene las estadísticas de resumen (summary).
- Las declaraciones plot **CDFPLOT**,**HISTOGRAM**, **PPPLOT**, **PROBPLOT** y **QQPLOT**, crea pantallas gráficas.
- **INSET** Mejora esas pantallas gráficas agregando una tabla de estadísticos en el gráfico.

PROCEDIMIENTO UNIVARIATE

- **CLASS** Para especificar una o dos variables que agrupan los datos en los niveles de clasificación. El análisis se hace para cada combinación de nivel.
- **BY** Usted lo especifica si desea un análisis por separado por cada grupo.
- **FREQ** Especifica la variable cuyos valores especifica la frecuencia para cada observación.
- **WEIGHT** Identifica una variable cuyos valores son los peso de cada observación en los cálculos estadísticos.
- **ID** Especifica una o más variables para identificar las observaciones extremas.

Ejercicios

Ejercicios: EL siguiente conjunto de datos contiene la presión sistólica y diastólica de 22 pacientes.

```
Title 'Ejemplo 1';
data Presion;
input PacienteID $ Sistolica Diastolica @@;
datalines;
CK 120 50  SS 96  60 FR 100 70
CP 120 75  BL 140 90 ES 120 70
CP 165 110 JI 110 40 MC 119 66
FC 125 76  RW 133 60 KD 108 54
DS 110 50  JW 130 80 BH 120 65
JW 134 80  SB 118 76 NS 122 78
GS 122 70  AB 122 78 EC 112 62
HH 122 82
;
run;
```

Ejercicios

```
proc univariate data=Presion;
run;

proc univariate data=Presion;
id pacienteID;
run;
```

Ejercicio: Simulación

```
data Normal1;
do i=1 to 100;
X= rannor(1)*10+100;
Y=1;
output;
end;
do i=1 to 100;
X= rannor(1)*5+150;
Y=2;
output;
end;
do i=1 to 100;
X= rannor(1)*8+200;
Y=3;
output;
end;
```

Ejercicio: Simulación

```
proc univariate data=NORMAL1 noprint;
class Y;
histogram X / NCOLS=1 nrows = 3;
inset mean std="Std Dev" / pos = ne format = 6.3;
run;
proc univariate data=NORMAL1 noprint;
class Y;
histogram X /  normal(percents=20 40 60 80 midpercents)
NCOLS=1 nrows = 3;
inset mean std="Std Dev" / pos = ne format = 6.3;
run;
```

Ejercicio: Simulación

```
proc univariate data=NORMAL1 noprint;
class Y;
histogram X /  normal(percents=20 40 60 80 midpercents)
Gamma
NCOLS=1 nrows = 3;
inset mean std="Std Dev" / pos = ne format = 6.3;
run;
proc univariate data=NORMAL1 noprint;
histogram X / kernel(c = 0.20 );
run;
```

Ejercicio: Simulación

```
proc univariate data=NORMAL1 noprint;
qqplot X ;
run;
proc univariate data=NORMAL1 noprint;
probplot X / normal(mu=150 sigma=10);
run;

proc univariate data=NORMAL1 noprint;
cdfplot X / normal;
inset normal(mu sigma);
run;
```

Ejercicio

- Usted tiene 2 bases de datos, que corresponde a los resultados de un test escolar en 2 años consecutivos. Las variables que usted encontrará en las datas son:
  - ID: Es la identificación del estudiante.
  - COLEGIO: Es la identificación del colegio.
  - Nivel prueba: Es el nivel educacional de las pruebas que va desde 1 hasta 11.
  - Pje: Es el puntaje del estudiante en la medición

Ejercicio

- Realice lo siguiente.
  1. Importe las Datas a SAS a una librería creada por usted.
  2. Una las Datas en una sola Base de Datos por ID.
  3. Elimine las inconsistencia de los Datos. Guárdela en una librería. Por ejemplo; El nivel del año 1 sea menor que la del año 2; Que el nivel del año 2 menos el año 1 tengan más de una diferencia de curso.
  4. Calcule el progreso de los estudiantes (puntaje año 2- puntaje año 1).
  5. Calcule el porcentaje de los estudiantes que cambiaron de colegio de un año a otro. Guárdelo en una Base de Datos que contenga 2 variables Nivel y cambio.

Ejercicio

6. Calcule el porcentaje de los estudiantes que cambiaron de colegio de un año a otro. Guárdelo en una Base de Datos que contenga 2 variables Nivel y cambio.
7. Identifique (imprima en pantalla) cual es el nivel que sufrió más cambio de estudiantes.
8. Ordena la DATA por colegio y nivel el segundo año y calcule el promedio de ambos puntajes.
9. Pegue la media que calculo en el ítem anterior en la data. Es decir para cada estudiante debe tener sus puntajes individuales de los test y además los de su grupo nivel-colegio.

Ejercicio

10. Ahora, calcule el promedio de cambio de estudiantes por nivel año y péguelo a la data anterior.
11. Finalmente la DATA final debe tener las variables ID, Colegio 2, nivel 2, cambio promedio del nivel colegio, puntaje año 1, puntaje año 2, promedio puntaje año 1 nivel colegio, promedio puntaje año 2 nivel colegio. En ese orden.
12. Realice un histograma por nivel, graficando la distribución normal y que aparezca la media y desviación estándar en el gráfico de la variable progreso creada anteriormente.

Ejercicio

13. Use Proc univariate para el del primer, segundo y tercer cuantil (en SAS Q1 Q2 Q3) por nivel. Guarde los resultados en la librería.
14. Realice un merge entre la base de datos que contiene progreso y los resultados de los cuantiles por nivel del año 2.
15. Crea una variable de clasificación, tal que:
    - 1 si progreso < que Q1
    - 2 si Q1 < progreso <Q2
    - 3 si Q2 < progreso <Q3
    - 4 si progreso > Q3

Ejercicio

16. Calcule el promedio del progreso por esta variable de clasificación y nivel.
17. Exporte a excel una data que contenga las variables nivel-clasificación-media del progreso, donde la hoja del excel se llame *medias-prog*.