SAS Workshop

Introduction to SAS Programming

Session II

Iowa State University

The Proc Step

A simple proc step:

PROC PRINT;

prints a *listing* of

- the most recently created data set
- all variables contained in the data set
- all observations in the data set

General form of the proc step:

PROC *proc-name options-list*;

Followed by

- procedure information statements
- variable attribute statements

The Proc Step

Examples of options:

DATA=NEW STDERR KENDALL

Examples of procedure information statements:

VAR … ;

BY … ;

CLASS … ;

MODEL … ;

OUTPUT … ;

Examples of variable attribute statements:

DROP … ;

FORMAT … ;

LABEL … ;

Examples of SAS Procedures

PROC MEANS MEAN VAR ;

VAR AGE INCOME ;

Other Options: DATA= , NOPRINT, MAXDEC=, N, MISS, MEAN, STD, MIN, MAX, RANGE, SUM, VAR, STDERR, CV, T, PRT etc.,

---

PROC CORR ;

VAR *HEIGHT WEIGHT*;

WITH *BP OXYGEN*;

Some Options: DATA= , PEARSON, SPEARMAN, BEST= , NOSIMPLE, NOPRINT, NOPROB etc.,

---

PROC UNIVARIATE ;

CLASS *COUNTY*;

VAR *ACREAGE RAINFALL*;

OUTPUT OUT= *NEW* MEAN= *AVE1 AVE2* VAR= *V1 V2* ;

Some Options: DATA= , NOPRINT, PLOT, NORMAL, PCTLDEF=,VARDEF=, ALPHA= , CIBASIC(TYPE= ALPHA=) , Muo=, TRIM=(TYPE= ALPHA=), etc.

Some Keywords for OUTPUT: N, NMISS, NOBS, MEAN, SUM, SD, VAR, SKEWNESS, KURTOSIS, MAX, MIN, RANGE, Q3, MEDIAN, Q1, P1, P5, P10, P90, P95, P99, etc.

LABEL and FORMAT Statements

LABEL *variable = 'label' ⋯* ;

e.g., LABEL PTS='AEROBIC POINTS'
PERCENT='% OF TOTAL FAT' ;

FORMAT *variables [format] ⋯* ;

e.g., FORMAT WEIGHT HEIGHT 8.2 INCOME DOLLAR12.2;

- These 2 statements can be used in both DATA and PROC steps
- In a DATA step, these statements will associate **labels** and **formats** permanently with the variable specified
- Also look-up LENGTH, INFORMAT, ATTRIB statements

Sample SAS Program B1

```
data sales;
infile ‘C:\Documents and Settings\mervyn\Desktop\sales.txt’;
input Region : $8. State $2. +1 Month monyy5. Headcnt Revenue Expenses;
format Month monyy5. Revenue dollar12.2;
run:

proc sort;
by Region State Month;
run;

proc print;
run;

proc print label;
        by Region State ;
        format Expenses dollar10.2 ;
        label State= State Month= Month Revenue='Sales Revenue'
              Expenses='Overhead Expenses';
 id Region State;
 sum Revenue Expenses;
 sumby Region;
 title ' Sales report by State and Region';
run;
```

Sample Data Set sales.txt

```
SOUTHERN FL JAN78 10 10000 8000
SOUTHERN FL FEB78 10 11000 8500
SOUTHERN FL MAR78  9 13500 9800
SOUTHERN GA JAN78  5  8000 2000
SOUTHERN GA FEB78  7  6000 1200
PLAINS NM MAR78  2   500 1350
NORTHERN MA MAR78  3  1000 1500
NORTHERN NY FEB78  4  2000 4000
NORTHERN NY MAR78  5  5000 6000
EASTERN NC JAN78 12 20000 9000
EASTERN NC FEB78 12 21000 8990
EASTERN NC MAR78 12 20500 9750
EASTERN VA JAN78 10 15000 7500
EASTERN VA FEB78 10 15500 7800
EASTERN VA MAR78 11 16600 8200
CENTRAL OH JAN78 13 21000 12000
CENTRAL OH FEB78 14 22000 13000
CENTRAL OH MAR78 14 22500 13200
CENTRAL MI JAN78 10 10000 8000
CENTRAL MI FEB78  9 11000 8200
CENTRAL MI MAR78 10 12000 8900
CENTRAL IL JAN78  4  6000 2000
CENTRAL IL FEB78  4  6100 2000
CENTRAL IL MAR78  4  6050 2100
```

Sample SAS Program B2

```
data biology;
input Id Sex $ Age Year Height Weight;
datalines;
7389   M  24  4  69.2  132.5
3945   F  19  2  58.5  112.0
4721   F  20  2  65.3   98.6
……..
……..
8472   M  21  2  76.5  205.1
6327   M  20  1  70.2  135.4
8472   F  20  4  66.8  142.6
4875   M  20  1  74.2  160.4
;

ods rtf file="progB2_out.rtf";
ods select BasicMeasures Quantiles TestsForNormality;
proc univariate data=biology Normal;
  var Height;
  title 'Biology class: Analysis of Height Distribution';
run;
ods rtf close;
```

Sample SAS Program B3

```
data biology;
input Id Sex $ Age Year Height Weight;
datalines;
7389   M  24  4  69.2  132.5
3945   F  19  2  58.5  112.0
4721   F  20  2  65.3   98.6
……..
……..
6327   M  20  1  70.2  135.4
8472   F  20  4  66.8  142.6
4875   M  20  1  74.2  160.4
;

proc means data=biology maxdec=3;
  class Year Sex;
  var Height Weight;
  output out=stats mean=Av_Ht   Av_wt  stderr=SE_Ht SE_wt;
run;

proc print data=stats;
  title 'Biology Class Data Set: Output Statement';
run;
```

Sample SAS Program B4

```
data world;
infile "demogr.txt";
input Country $20. Birthrat Deathrat Inf_mort Life_exp Popurban
          Perc_gnp Lev_tech Civillib;
run;

title1 c = magenta f = triplex ' Normal Probability Plot ';
symbol1 c = salmon v = star i = none;

proc univariate data=world;
  var Inf_mort;
  qqplot Inf_mort/normal(mu=est sigma=est) ca=blue
          ct= red  vm=9  vaxislabel='Infant Mortality' pctlminor;
  inset mean std /format=8.2 header='Reference Line'
                          refpoint=bl position=(75,10);
run;
```

Sample SAS Program B4 (Continued)

```
title 'Distribution of Infant Mortality';
axis1 label=(a=90 r=0);

ods select ParameterEstimates GoodnessOfFit MyHist;
proc univariate data=world;
  var Inf_mort;
  histogram Inf_mort/ midpoints=10 to 160 by 20 lognormal
            vaxis = axis1 name  = 'MyHist';
  inset n mean (5.3) std='Std Dev' (5.3) skewness (5.3)/
       pos = ne  header = 'Summary Statistics';
run;
```

Sample SAS Program B5

```
data algae;
 input Clumps $ Count;
 label Clumps  ='Cell clumps/field'
       Count  ='Number of fields';
 datalines;
 0  6
 1 23
 2 29
 3 31
 4 27
 5 13
 6  8
 >7 13
;

ods rtf file="progB5_out.rtf";
proc freq data=algae order=data;
   weight Count;
   tables Clumps/ nocum testp=(3.7 12.2 20.1 22.1 18.2 12.0 6.6 5.1);
   title 'Fitting a Poisson to cell clump counts';
run;
ods rtf close;
```

Sample SAS Program B6

```
data muscle;
input x y;
label x='Age' y='Muscle Mass';
datalines;
71 82
64 91
…….
49 105
78 77
;
ods listing close;
ods pdf file="progB6_out.pdf";
ods graphics on;
proc reg data=muscle plots(only)=(diagnostics qq residualbypredicted fit residuals);
  model y = x/r;
  title 'Simple Linear Regression Analysis of Muscle Mass Data';
run;
ods graphics off;
ods pdf close;
ods listing;
```

Sample SAS Program B7

```
proc import out= work.air
            datafile= "air_pollution.xls"
            dbms=xls replace;
            getnames=yes;
run;

proc print data=air;
run;

ods listing close;
ods rtf file="progB7_out.rtf" style=statistical;

proc sgscatter data=air;
  title "Scatterplot Matrix for Air Pollution Data";
  matrix SO2--PrecipDays;
run;
proc reg corr data=air;
 model SO2 =  AvTemp--PrecipDays/clb vif;
 title 'Model fitted with all explanatory variables';
run;
ods rtf close;
ods listing;
```

Sample SAS Program B7 (Continued)

```
data names;
infile "air_pollution_names.txt" truncover;
input City CityName $14.;
run;

proc print data=names;
run;

data mylib.pollution;
 merge air names;
 by City;
run;

proc print data=mylib.pollution;
run;
```

Sample SAS Program B8

```
data new;
set temp.pollution;
if _N_= 31 then delete;
run;

ods listing close;
ods pdf file="progB8_out.pdf";

ods graphics on;
proc reg data=new plots(only)=(criteria cp(label));
 model SO2 =  AvTemp--PrecipDays/selection=rsquare start=2 stop=4
                                                    best=4 cp sse mse;
 title 'Models fitted with all explanatory variables (with Obs#31 deleted)';
run;
ods graphics off;
ods pdf close;
ods listing;
```