Understanding some SAS programs

In today’s class I will go over some SAS programs and try to explain what they do. We will also encounter some new PROCS - especially for graphing. These will be very fundamental SAS plots in that they do not employ the SAS GRAPHS package. These plots use what are called character plotting and enable you to see plots on any text editor. Therefore they are highly portable. On the downside, these plots do not appear as “professional quality” to be included in your publications. For that you will need to learn SAS graphics - I will not be able to cover that in this course. I hope all of you are now familiar with running a SAS program on saturn, using the pico editor. Now let us try to understand a bit more about SAS.

1 A complete SAS program

We now present the following program in its entirety, which you may write and save as example2.sas. Then you may run it from the command prompt typing (DO NOT FORGET the space between sas and the filename):

```
sas example2.sas
```

Below is the complete program.

```
options ls = 80 nodate;
data htwt;
input gender $ height weight;
datalines;
M 68.5 155
F 61.2 99
F 63.0 115
M 70.0 205
M 68.6 170
F 65.1 125
M 72.4 220
```
run;
proc univariate data = htwt normal plot;
run;

proc freq data = htwt;
title 'Using Proc Freq to compute frequencies';
tables gender;
run;
proc chart data = htwt;
vbar gender;
run;
proc chart data = htwt;
hbar gender;
run;
proc chart data = htwt;
title 'Distribution of heights by levels';
vbar height / levels=6;
run;
proc chart data = htwt;
title 'Distribution of heights by midpoint rule';
vbar height / midpoints = 50 to 80 by 10;
run;
proc plot data=htwt;
title 'Simple scatter plot';
plot weight*height;
run;
proc plot data=htwt;
title 'Scatter plot by gender';
plot weight*height = gender;
run;
2 Dissecting the program

Let us now dissect the program.

```sas
options ls = 80 nodate;
data htwt;
input gender $ height weight;
datalines;
M 68.5 155
F 61.2 99
F 63.0 115
M 70.0 205
M 68.6 170
F 65.1 125
M 72.4 220
;
run;
```
The top line sets line-size options and suppresses printing of dates. Next we proceed to the creation of a SAS data set. The $ sign after gender indicates the first variable is a character. We have seen this step in the lab in all its gory details.

```sas
proc univariate data = htwt normal plot;
run;
```

Now we invoke procedure `univariate` to perform a simple descriptive statistics on the numeric variables in our dataset. Adding the `normal` and `plot` options directs this procedure to supply basic plots like stem-leaf, box-plots and normal quantile plots. Explanations of some of the relevant portions of the output are given below:

N: Number of non-missing observations.

Sum: Wgts: Sum of Weights (if WEIGHT statement is used)

Mean: Arithmetic mean

Sum: Sum of the scores.

Std Dev: Standard deviation

Variance: Variance (square of the standard deviation)

Skewness: Measure of the symmetry or asymmetry of the distribution.

Kurtosis: Measure of the flatness of the distribution

USS: Uncorrected sum of squares (Each score is squared and the squares are added together)

CSS: Corrected sum of squares (Sum of squares about the mean, usually more useful than the median)

CV: Coefficient of variation

Std Mean: Standard error of the mean (the standard deviation divided by the square root of the mean).
Our next procedure is called `proc freq` and it is used to dissect data by frequency. It helps in analyzing data by different groups. In our case the following code will actually provide separate descriptions for females and males:

```plaintext
proc freq data = htwt;
title 'Using Proc Freq to compute frequencies';
tables gender;
run;
```

Next we see some plotting procedures. The following four blocks of code all produce character plots. The first two produce vertical and horizontal bar charts. The next two are two different ways of producing histograms.

```plaintext
proc chart data = htwt;
vbar gender;
run;
proc chart data = htwt;
hbar gender;
run;
proc chart data = htwt;
title 'Distribution of heights by levels';
vbar height / levels=6;
run;
proc chart data = htwt;
title 'Distribution of heights by midpoint rule';
vbar height / midpoints = 50 to 80 by 10;
run;
```

Next we look at bivariate plots. As we disussed in the previous lecture, these are extremely useful in understanding the relationships between two variables. The following represent different ways of producing these plots. The key procedure is `proc plot`.

```plaintext
proc plot data=htwt;
title 'Simple scatter plot';
```

5
plot weight*height;
run;

proc plot data=htwt;
  title 'Scatter plot by gender';
  plot weight*height = gender;
run;

Is it possible to produce different scatter graphs for the two different groups (males and females)?
Certainly - First sort the data by gender. Note: proc sort sorts the dataset as directed, but does not produce any output.

proc sort data = htwt;
by gender;
run;

Next invoke proc plot in the following way.

proc plot data=htwt;
  title 'Separate plots';
  by gender;
  plot weight*height;
run;