

Johnson Chapter 1 Exercise 8 – Hints

A Scatterplot is obtained using either PROC PLOT or PROC GPLOT. PROC PLOT produces a low resolution scatterplot that is displayed in the Listing tab. PROC GPLOT produces a high resolution scatterplot that is scalable. Basic syntax is:

```
Proc GPlot Data=A;
  Plot Y*X;
Run;
```

Y is the variable plotted on the vertical axis.
X is the variable plotted on the horizontal axis.

PROC MEANS and PROC CORR can be used to obtain mean vectors, variance-covariance matrices, and correlation matrices.

The use of the BY statement in PROC MEANS and PROC CORR will produce separate analyses for each level of the variable specified in the BY statement. For example, if GROUP is a categorical variable with 4 levels (say 1, 2, 3, 4), then the following code will produce the mean of the variable Y in each of the 4 groups:

```
Proc Sort Data=A;
  By Group;
Run;

Proc Means Data=A Mean;
  Var Y;
  By Group;
Run;
```

Note that, as above, before using the BY statement with PROC MEANS, you should first sort the data by the variable used in the BY statement.

PROC STDIZE can be used to produce standardized z-scores. Let the SAS data set A contain the (unstandardized) variable Y. Consider the following code:

```
Data B;
  Set A;
  Z=Y;
Run;

Proc Stdize Data=B Out=C Method=Std;
  Var Z;
Run;
```

The DATA step creates a new SAS data set named B, which contains everything in the data set A (in this case, Y), and a new variable named Z. At this stage, Y and Z are identical. PROC STDIZE takes the variable Z from data set B, standardizes it by subtracting the mean and dividing by the standard deviation, and output the result to a new data set C. The SAS data set C now contains the original variable Y and the standardized variable Z. You can print these standardized scores using PROC PRINT.

```
Proc Print Data=C;  
  Var Z;  
Run;
```