

Data learning 4 : Introduction to SAS software

Enter the SAS Software. You have 3 windows, **Explorer** that allows you to view and manage your files, **Log** that displays messages about your SAS session, for example error messages and **Editor** that is used to edit and submit the SAS programs. There are two windows more, **Output** that will appear to display the outputs of the SAS programs and **Result** that can be viewed instead of **Explorer** and that help you to navigate among the results.

Exercise 1 : Create, import, export data sets

1. Click in **Editor** and using **File/Open** edit the file called "Exercice1.sas". It contains 4 lines of code and a small table of variables. In **Explorer**, click on **Libraries/Work**. Coming back to **Editor**, highlight all the lines of the program and submit them using the running-man button. Read the message in **Log**. Notice that a table called "Socioeconomics" is now display in **Explorer**. A file with this name has been created. Click on it to see the file but DON'T FORGET TO CLOSE THE VIEWTABLE AFTER.
2. To import a data file and transform it in a SAS-data set, use **File/Import Data** and choose the type **Delimited file (*.*)** to import the file "Cancer.txt" with the name Cancer1 in the library **Work**. Have a look on it but DON'T FORGET TO CLOSE THE VIEWTABLE AFTER.
3. In **Explorer**, at the same level as **Work** there is 3 other libraries. Click on the one called "Citywk" (weekly financial datas). Describe the content of the file (how many variables, how many individuals,). Using **File/Export Data**, create a new file "Citywk.xls" on your disk and open it with Excel to see if every thing is OK.

Exercise 2 : Performe a PCA with Analyst. To enter the **Analyst** environment, choose **Solution/Analysis/Analyst**. Then select **Tools/SampleData/Protein** and **File/Open by SasName/OK** to open in SAS the file that was studied in our first lecture. To perform a PCA, select **Statistics/Multivariate/Principal Components**. Highlight all the variables and click on the **Variables** button to select all the variables for the analysis. To request a scree plot, click the **Plot** button and **create scree plot**. To request a plot on the subspace of the two first principal components, click **Plot** again but choose the windows **ComponentPlot** and then **Create component Plots**. Choose the type **Enhanced** and highlight "Country" before to click on the button **id** and **OK**. Interpret all the outputs you have produce.

Notice that the code SAS that has been used by the software to perform this PCA and plot the results is available.

Exercise 3 : Perform a PCA without the help of Analyst. If you have just to compute the coordinates of the principal components, it can be faster to directly use the procedure **FACTOR**. For example, to compute PC of the data **SocioEconomics** we create before, type in the **Editor** the following lines : `proc Factor data=SocioEconomics simple corr ; run ;`

Submit this program and try to explain the results you obtain. To see how this exemple is explained in the "Doc on line", edit this doc, choose the last item **SAS/STAT**, than the **SAS/STAT User's Guide** and finally the **FACTOR Procedure, example 26.1**.

Exercise 4 : Use of the SAS Doc on line. Still in the Doc on line, see the **PRINTCOMP Procedure, Example 52.1**. Highlight the code SAS (between "Crime;" until "run;") and copy and past it in a new **Editor Windows**. Submit and see the results.

Exercise 5 : An other PCA. Using the environment **Analyst**, perform a PCA of the set **Citywk** that you can find in the **Sashelp** library (choose the variables you want for the analysis).

Exercise 6 : An example of Correspondance Analysis. With the help of the notes of the lecture 2, perform a correspondance analysis of the data set "PhD.txt" : first import the data file in SAS, then use the procedure **CORRESP** and also `%plotit` to create the picture.

Exercise 7 : Another example of Correspondance Analysis. Do the same for the data set "US-pop.txt".

Exercise 8 : Example of Cluster Analysis. With the help of the notes of the lecture 2, perform a cluster analysis of the data set "Cities.txt" taht show the flying Mileages between 10 American cities : first import the data file in SAS, then perform a HA using the procedure **CLUSTER**. Plot the dendrogram. How many clusters would you keep ?