Automate a routine analysis, example of Principal Component Analysis, in XLSTAT

Automation.zip

Dataset for automating a routine analysis

Two Excel workbooks with both the data and the results can be downloaded by clicking here.

The data used is the process measurements of food samples.

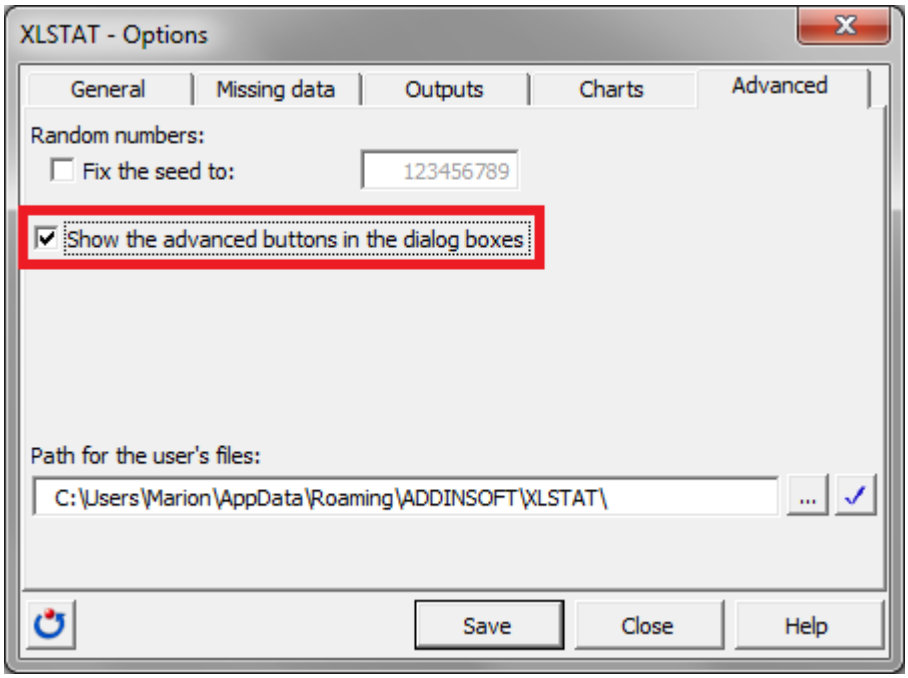
Creating the VBA codes to be reused

We are going to create a Principal component analysis template on one dataset and use it on the second.

Generating the code to automate a routine analysis

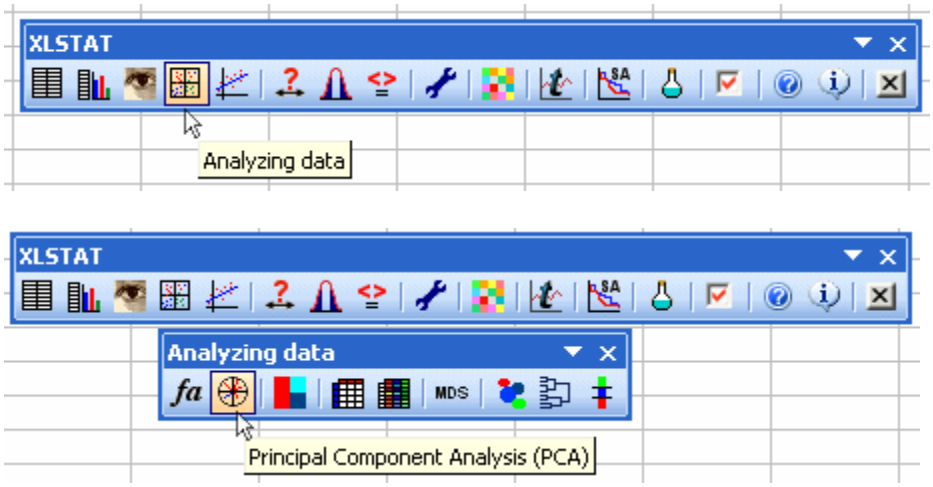
Open the first file Automation_1.xls

Once XLSTAT-Pro is activated, go to the menu **Options** and in the tab **Advanced** enable the option **Show the advanced buttons in the dialog boxes**.



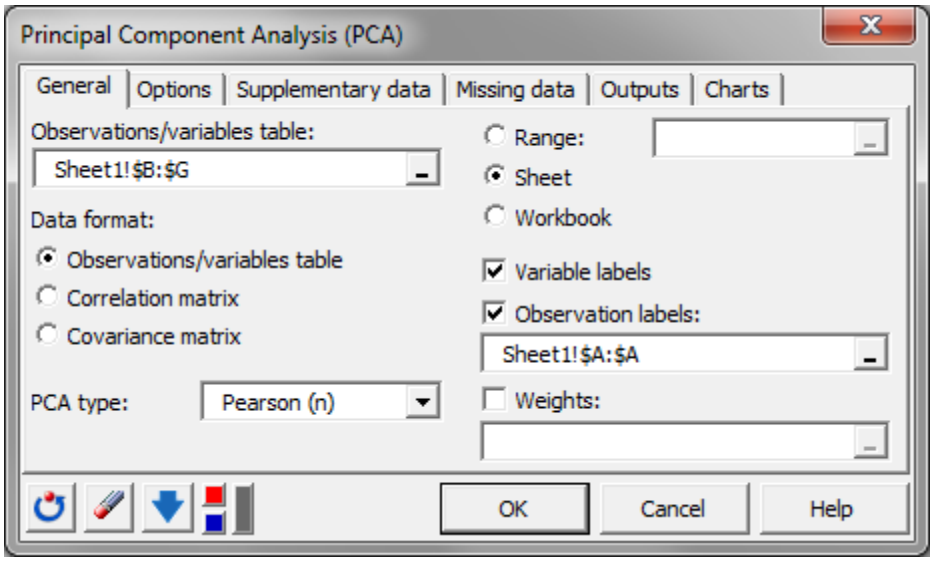
The next step of the automation procedure is to set up your statistical analysis.

Select the **XLSTAT / Analyzing data / Principal components analysis** command, or click on the corresponding button of the **Analyzing Data** toolbar (see below).

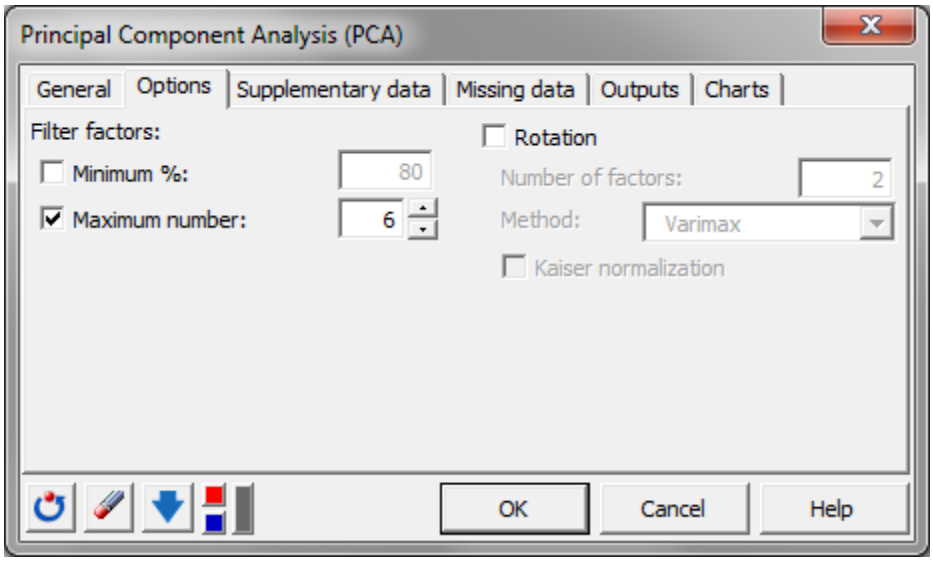


In the **General** tab, set the following:

- **Observations/variables table**: Columns B to G
- **Data format**: Observations/variables table
- **PCA type**: Pearson (n)
- **Variable labels**: enabled
- **Observation labels**: ticked and select the column A for the sample name
- **Sheet**: chosen to display the results in a new sheet

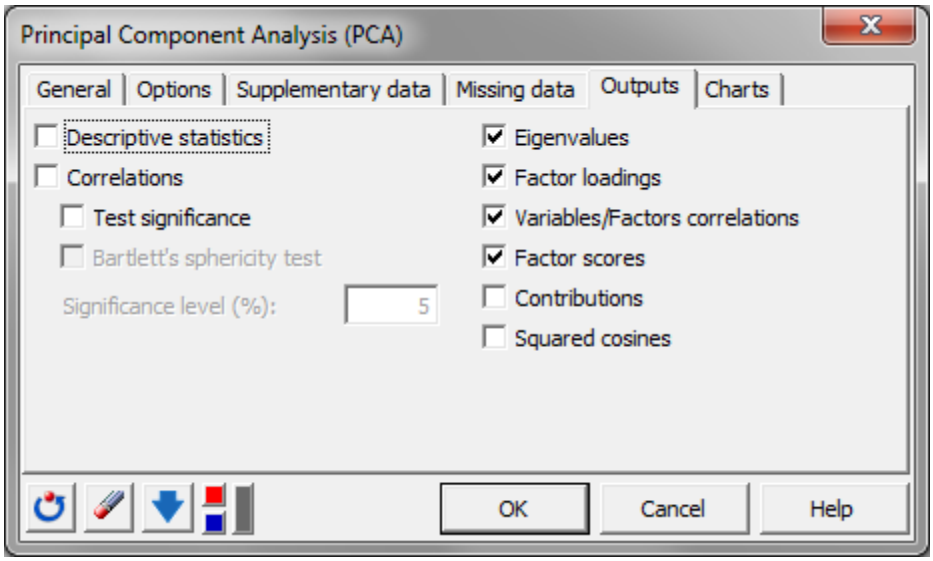


Go to the next tab **Options**. For the option **Filter factors**, choose **Maximum number** and set the value to six. This way all the components will be calculated.



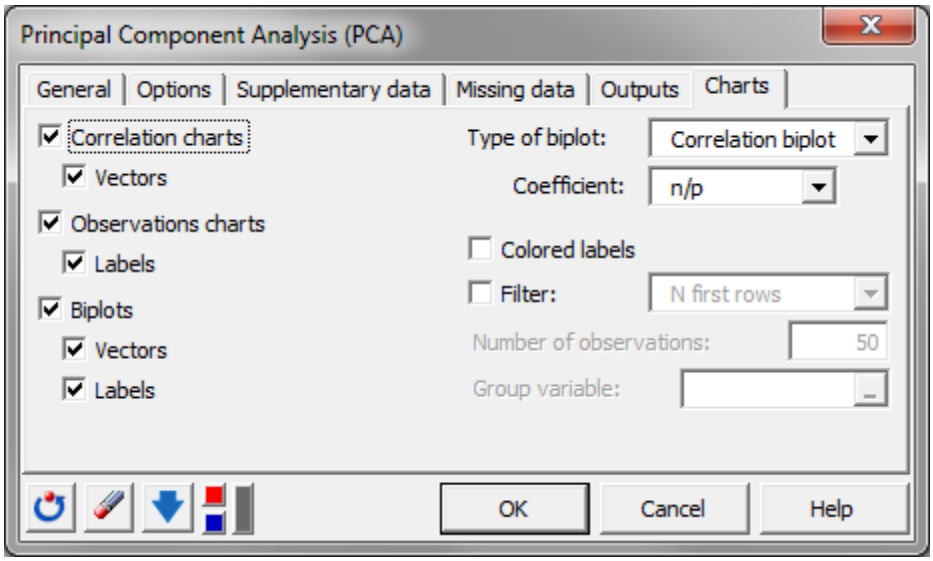
Go to the tab **Outputs**. Here we want to get a synthetic report so we will only select the following:

- Eigenvalues,
- Factor Loadings,
- Variables/Factors correlations,
- Factor scores.



Finally we are going to use all three plots that can be selected in the **Charts** tab:

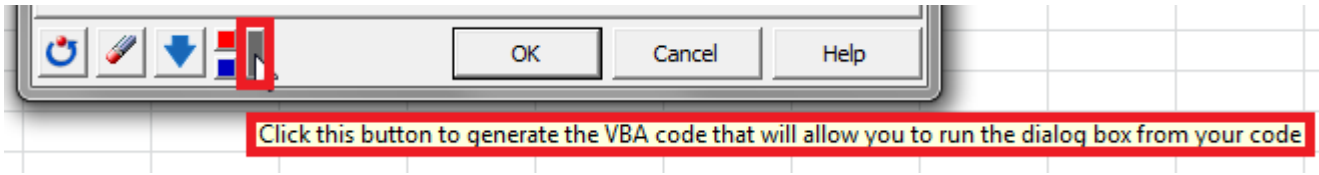
- Correlation charts
- Observations charts
- Biplots



Now we have specified all the setting we will save the code to be reused.

Generate the VBA code to be reused

Click on the grey button at the bottom left of the dialog box: **Click this button to generate the VBA code that will allow you to run the dialog box from your code**.



Once you have pressed the button a Notepad document will appear containing the VBA code. Save the code under a name that is easy for you to remember, for example in this case we use "VBA-PCA-recipe1".

VBA-PCA-recipe1 - Bloc-notes

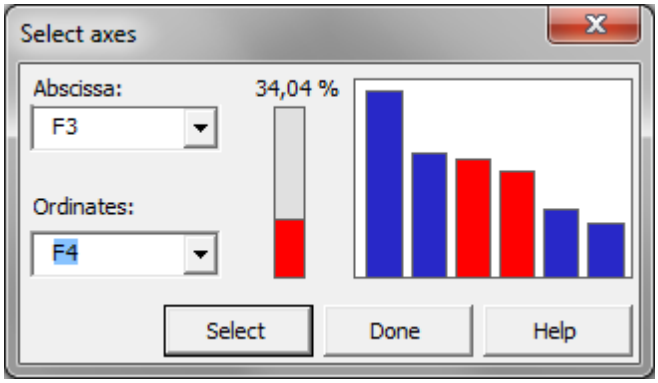
Fichier Edition Format Affichage ?

```
Sub RunMeOnce()

        Call ActiveWorkbook.VBProject.References.AddFromFile("C:\Program Files (x86)\Addinsoft
\XLSTAT2010\XLSTAT-MCA.dll")

End Sub

Sub MySub()

        Call LoadRunPCA(Range("Sheet1!$B:$G"), ObsLabels:=Range("Sheet1!$A:$A"), PCAType:=2,
NoScreenUpdating:=True, SettingsFile:="C:\Users\Marion\AppData\Roaming\ADDINSOFT\XLSTAT
\Form22.txt")

End Sub
```

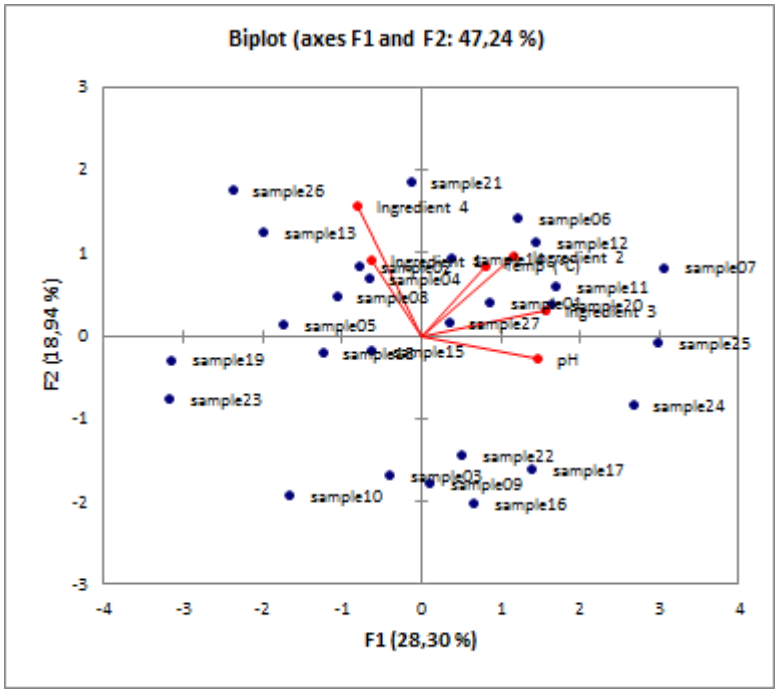
Results of the analysis

Click on **OK** to launch the analysis.

Now choose the plot for the axes F1 and F2 by clicking **Select**, then change the selection to **Abscissa** F3 and **Ordinates** F4. Once you have completed this click again on **Select** and then press **Done**.



Have a look at the biplot.

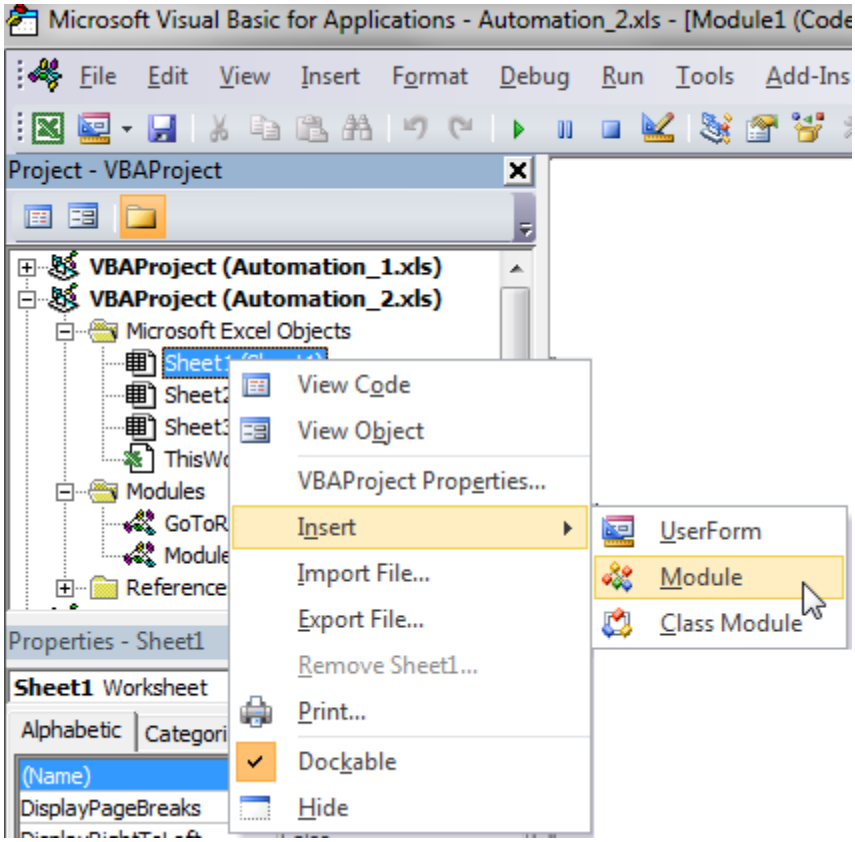


This process is usually stable so we can expect little variation. You can see that all the samples are centered tidily around the middle of the plot.

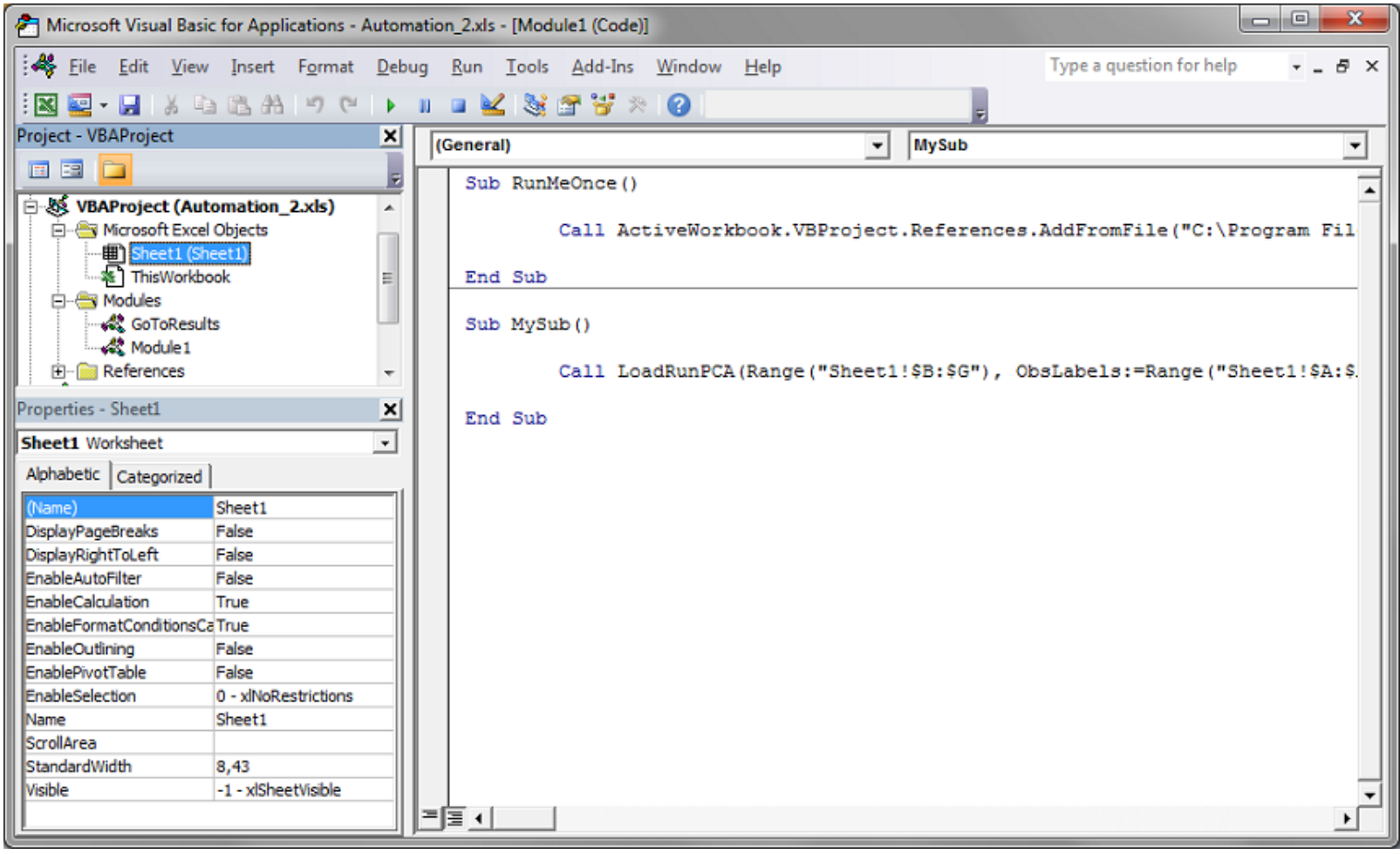
Reusing the VBA code

Now open the second file Automation_2.xls

Press **Alt+F11** together in order to launch the Visual Basic Application. Then select **Sheet1** in the folder **VBAProject(Automation_2.xls)** and finally right click and opt for the action **Insert / Module**

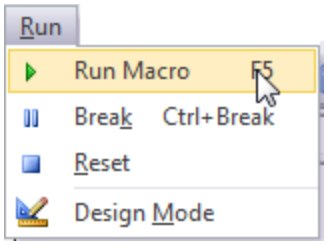


The next step is to copy and paste the code contained in the Notepad file into this module.

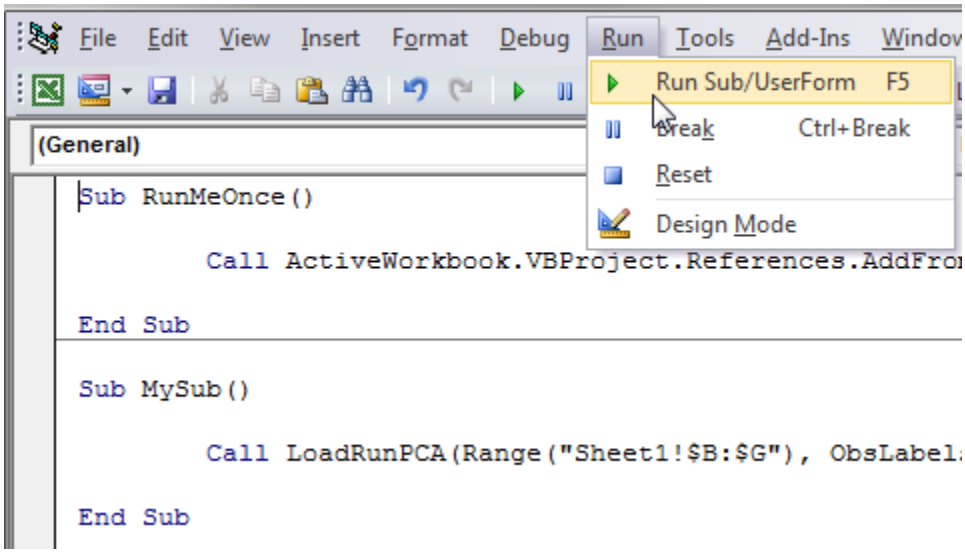


At this step you can add more codes to enable the program to perform other actions.

Go to the menu **Run / Run Macro** located in the menu bar.



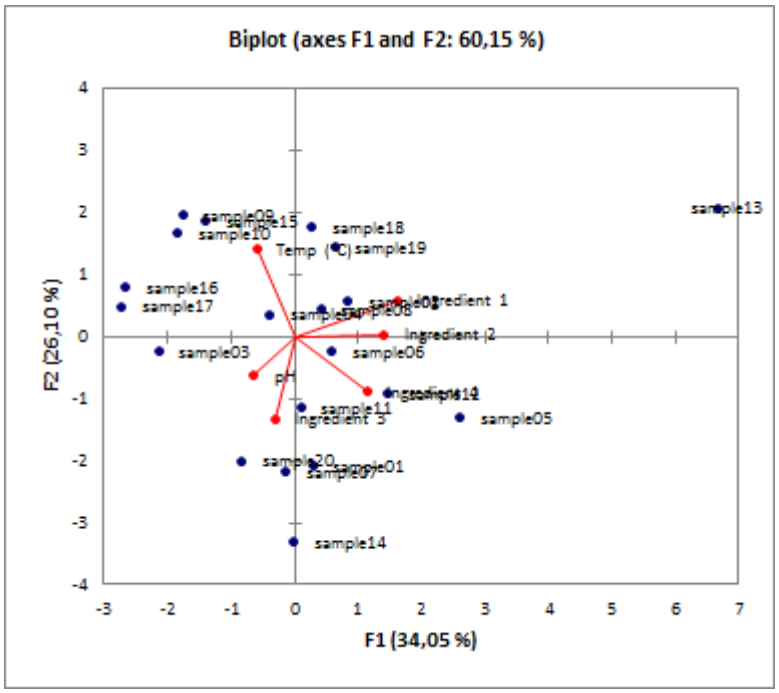
Then you need to run first the macro called "RunMeOnce". This will make a link between the file and the XLSTAT project where the code is stored. Select it in the list and click on **Run**.



When this has been completed, run the second macro called "MySub". Return to the menu **Run / Run Macro** and this time select the macro "MySub" before pressing the button **Run**.

This will in turn execute the code in question and you now have a sheet "PCA" containing the results.

Now if we look at the biplot of the second analysis we notice that this time one of the samples seems to be further away than the other samples. Sample 13 may be an outlier.



Have a look at this video to see a demonstration on how to automate data analysis with XLSTAT software.

http://www.youtube.com/watch?v=iF2hmxn2W2o&feature=player_embedded