This tutorial contains navigation buttons that enable you to move throughout the tutorial.

Please use the navigation buttons and not the page up/page down or arrow keys to navigate through the tutorials.

This is the 'Next' button. It takes you to the next frame or stop point.

This is the 'Previous' button. It takes you to the previous frame or stop point.

This is the 'Go to frame' button. It takes you to a specified frame.

This is the 'Go to URL' button. It takes you to a website link.

Press the 'Next' button below to start this tutorial.
This tutorial will show how to create a PCA model and then project new data into the model.

For this we will use the 'Run PCA' function to create and save the model and then the 'PCA Modelling' function to project the new data into the model.

Before doing this procedure you will need to have a data set for building the model and a separate data set that you want to project into the model.

Each set of data needs to have its own list of files and samplenames.
First we need to build a PCA model.
From the 'MVA' menu choose -> 'Run PCA'
Data Selection Panel

These are the main input data that will be used in further analysis unless you specify otherwise. Use the drop down menus to select the data and information you want to use in your analysis.

Name of Data Matrix: ndataset
Name of Variable Matrix: exactmass
Name of Filename Matrix: filenames
Name of Totalcounts Matrix: totalcounts
Name of Samplenames Matrix: samplenames

Input Information

Make sure the data selected above is the correct set of data is loaded with the correct filename and samplename list.

Load Selected Data

Data: None
Samples: None
Variables: None

Scaling Method
Choose one

Run PCA

Show Eigenvalue Scree Plot

PCA Summary
PCA %Var %VarTotal

Close Panel
Data Selection Panel

These are the main input data that will be used in further analysis unless you specify otherwise. Use the drop down menus to select the data and information you want to use in your analysis.

Input Information

Make sure the data selected above is the data you want to analyze and then press the "Load Selected Data" button.

Press the 'Load Selected Data' button.

PCA Panel

Scaling Method

Choose one

Run PCA

Show Eigenvalue Scree Plot

PCA Summary

PC# %Var %VarTotal

Close Panel
Data Selection Panel

These are the main input data that will be used in further analysis unless you specify otherwise. Use the drop down menus to select the data and information you want to use in your analysis.

<table>
<thead>
<tr>
<th>Name of Data Matrix</th>
<th>Name of Variable Matrix</th>
<th>Name of Filename Matrix</th>
<th>Name of Totalcounts Matrix</th>
<th>Name of Samplenames Matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>ndatass</td>
<td>exactmass</td>
<td>filenames</td>
<td>totalcounts</td>
<td>samplenames</td>
</tr>
</tbody>
</table>

Input Information

Make sure the data selected above is the data you want to analyze and then press the "Load Selected Data" button.

PCA Panel

Choose the scaling/preprocessing you would like to use.

NOTE: The choice you make here is automatically applied to the new data when it is projected into the PCA model.
Data Selection Panel

These are the main input data that will be used in further analysis unless otherwise specified. Use the drop-down menus to select the data and information you want to use in your analysis.

Name of Data Matrix: ndata
Name of Variable Matrix: exactmass
Name of Filename Matrix: filenames
Name of Totalcounts Matrix: totalcounts
Name of Samplenames Matrix: samplenames

Input Information

Make sure the data selected above is the data you want to analyze and then press the "Load Selected Data" button.

Load Selected Data

Data: ndata
Samples: samplenames
Variables: exactmass

Scaling Method
Mean Center

Press the 'Run PCA' button.

PCA Summary

PCA %Var %VarTotal

Close Panel
When creating a PCA model, the number of PCs kept in the model is very important. Selecting too few PCs will create a model that may not have good predictive ability, selecting too many PCs will create a model that is too restrictive to the calibration set and that will be modeling noise.

For this data set looking at the % variance it looks like the numbers stop changing significantly around PC 7 or 8.
Let's see what the scree plot looks like.

Press the 'Show Eigenvalue Scree Plot' button.
It looks like there are two inflection points, one at 6 and one at 8.

Let's look at the scores and loadings plots and see what PCs 6 through 8 look like.
First we'll look at PC 6.
It looks like PC5 separates the 16 and 18 samples. So there is some information here that looks useful.
PC7 looks like it is capturing some scatter in sample 6, and maybe something in sample 18.
Data Selection Panel

These are the main input data that will be used in further analysis unless you specify otherwise. Use the drop-down menus to select the data and information you want to use in your analysis.

Input Information
Make sure the data selected above is the data you want to analyze and then press the "Load Selected Data" button.

Load Selected Data

PCA Panel

Plot PCA Scores and Loadings
These plots are for a quick check of the results and cannot be saved.
Loading plots default to m/z vs PC#.

Save PCA Data to Workspace

PC8 definitely looks like it is mainly noise in the data.
Data Selection Panel

These are the main input data that will be used in further analysis unless you specify otherwise. Use the drop-down menus to select the data and information you want to use in your analysis.

Input Information

Make sure the data selected above is the data you want to analyze and then press the "Load Selected Data" button.

PCA Panel

Plot PCA Scores and Loadings

These plots are for a quick check of the results and cannot be saved.

Loading plots default to m/z vs PC#.

Show Eigenvalue Scree Plot

PCA Summary

<table>
<thead>
<tr>
<th>PCA</th>
<th>%Var</th>
<th>%Var Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>52.3</td>
<td>52.3</td>
</tr>
<tr>
<td>2</td>
<td>20.6</td>
<td>73</td>
</tr>
<tr>
<td>3</td>
<td>13.7</td>
<td>86.7</td>
</tr>
<tr>
<td>4</td>
<td>7.1</td>
<td>93.8</td>
</tr>
<tr>
<td>5</td>
<td>3.5</td>
<td>97.2</td>
</tr>
<tr>
<td>6</td>
<td>1.5</td>
<td>98.8</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>99.7</td>
</tr>
<tr>
<td>8</td>
<td>0.1</td>
<td>99.8</td>
</tr>
<tr>
<td>9</td>
<td>0.1</td>
<td>99.9</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>99.9</td>
</tr>
</tbody>
</table>

Note how PC3 showed scatter while there is at least some pattern in PC7.
Data Selection Panel
These are the main input data that will be used in further analysis unless you specify otherwise. Use the drop-down menus to select the data and information you want to use in your analysis.

Name of Data Matrix: ndass
Name of Variable Matrix: exactmass
Name of Filename Matrix: filenames
Name of TotalCounts Matrix: totalcounts
Name of SampleNames Matrix: samplenames

Input Information
Make sure the data selected above is the data you want to analyze and then press the "Load Selected Data" button.

Load Selected Data
Data: ndass
Samples: samplenames
Variables: exactmass

Scaling Method
Mean Center

Run PCA

Show Eigenvalue Scree Plot

PCA Panel
Plot PCA Scores and Loadings
These plots are for a quick check of the results and cannot be saved.

While there may be some question about whether to keep 6 or 7 PCs for this model, for the sake of this tutorial we will keep all PCs up to where there is just noise in the scores.

By this choice I am not suggesting a selection criterion. I suggest you search the literature for more information on selecting the number of PCs for a model.
Data Selection Panel

These are the main input data that will be used in further analysis unless you specify otherwise. Use the drop-down menus to select the data and information you want to use in your analysis.

Name of Data Matrix: ndatas
Name of Variable Matrix: exactmass
Name of Filename Matrix: filenames
Name of Totalcounts Matrix: totalcounts
Name of Samenames Matrix: samenames

Input Information

Make sure the data selected above is the data you want to analyze and then press the "Load Selected Data" button.

Load Selected Data

Scaling Method:
Mean Center

Run PCA

Show Eigenvalue Scre Plot

PCA Summary

<table>
<thead>
<tr>
<th>PCA</th>
<th>% Var</th>
<th>% Var total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>52.3</td>
<td>52.3</td>
</tr>
<tr>
<td>2</td>
<td>20.6</td>
<td>73</td>
</tr>
<tr>
<td>3</td>
<td>13.7</td>
<td>86.7</td>
</tr>
<tr>
<td>4</td>
<td>7.1</td>
<td>93.8</td>
</tr>
<tr>
<td>5</td>
<td>3.5</td>
<td>97.2</td>
</tr>
<tr>
<td>6</td>
<td>1.5</td>
<td>98.8</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>99.7</td>
</tr>
<tr>
<td>8</td>
<td>0.1</td>
<td>99.8</td>
</tr>
<tr>
<td>9</td>
<td>0.1</td>
<td>99.9</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>100</td>
</tr>
</tbody>
</table>

Plot PCA Scores and Loadings

These plots are for a quick check of the results and cannot be saved. Loading plots default to m/z vs PC.

Numbers of PCs to keep: 7

Now enter names for the scores, loadings, % variance, and model. These should be unique names that have not been used for other variables in the Matlab workspace.

Scores output name:

Loadings output name:

Percent variance output name:

Model output name:

Save To Workspace

Close Panel
Now we will use the model we just created and project new data into it.
From the 'MVA' menu choose -> 'PCA Modelling'
Data Selection Panel
These are the main input data that will be used in further analysis unless you specify otherwise. Use the drop down menus to select the data and information you want to use in your analysis.

Name of Data Matrix: ndataset
Name of Variable Matrix: exactmass
Name of Filename Matrix: filenames
Name of Totalcounts Matrix: totalcounts
Name of Samplenames Matrix: samplenames

PCA Modelling Panel
Choose a Model to Load
Select Model

Select the data you want to project

Model Information

Original Data Set
Data: None
Variables: None
Samples: None

Model Data
Scores: None
Loads: None
% Variance: None
Scaling Choice: None

Save Model Data to Workspace
Name for New Data Scores
Name for combined scores (new and model)
Save New Scores
Save Combined Scores

Project Data Into Model

Close Panel
Data Selection Panel
These are the main input data that will be used in further analysis unless you specify otherwise. Use the drop down menus to select the data and information you want to use in your analysis.

Name of Data Matrix
- ndataset

Name of Variable Matrix
- exactmass

Name of Filename Matrix
- filenames

Name of Totalcounts Matrix
- totalcounts

Name of Samplenames Matrix
- samplenames

PCA Modelling Panel

Choose a Model to Load
- pcamodelf

Select the data you want to project into the model, and corresponding samplenames from the 'Raw Data Selection Panel'.

Save Model Data to Workspace

Model Information
- Original Data Set
  - Data: ndataset
  - Variables: exactmass
  - Samples: samplenames

Model Data
- Scores: pccscores
- Loads: pcacore
- Variance: pcavariance
- Scaling Choice: Mean Center

The model keeps a reference to what data was used to create the model. That way you can remember exactly what data was used so you can go back and adjust the model if you need to.

This information is listed here in the 'Original Data Set' box.

Save New Scores
Save Combined Scores

Close Panel
Data Selection Panel
These are the main input data that will be used in further analysis unless you specify otherwise. Use the drop-down menus to select the data and information you want to use in your analysis.

Name of Data Matrix: ndataset
Name of Variable Matrix: exactmass
Name of Filename Matrix: filenames
Name of Totalcounts Matrix: totalcounts
Name of Samplenames Matrix: samples

PCA Modelling Panel

Choose a Model to Load
pcamodel

Select the data you want to project into the model, and corresponding sample names from the 'Raw Data Selection Panel'

Save Model Data to Workspace
Name for New Data Scores
Name for combined scores (new and model)

Model Information
Original Data Set
Data: ndataset
Variables: exactmass
Samples: samples

Model Data
Scores: pcascores
Loads: pcaloads
% Variance: pcvairance
Scaling Choice: Mean Center

The 'Model Data' box lists the names of the PCA results that are stored in the model and the scaling/pre-processing methods used in the model.
After loading the model, select the new data that you want to project into the model from the 'Data Selection Panel'.

You need to select the data and samenames. Here I also select the filenames though they are not required for the function.
## Data Selection Panel

These are the main input data that will be used in further analysis unless you specify otherwise. Use the drop down menus to select the data and information you want to use in your analysis.

<table>
<thead>
<tr>
<th>Name of Data Matrix</th>
<th>Name of Variable Matrix</th>
<th>Name of Filename Matrix</th>
<th>Name of Totalcounts Matrix</th>
<th>Name of Samplenames Matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>intestdataset</td>
<td>exactmass</td>
<td>testfiles</td>
<td>totalcounts</td>
<td>testsamples</td>
</tr>
</tbody>
</table>

## PCA Modelling Panel

**Choose a Model to Load**

- pcamodel

**Model Information**

- **Original Data Set**
  - Data: intestdataset
  - Variables: exactmass
  - Samples: samplenames

**Select the data you want to project into the model, and corresponding samplenames from the 'Raw Data Selection Panel'**

- Load Selected Data

**Save Model Data to Workspace**

- Name for New Data Scores
- Name for combined scores (new and model)
- Save Combined Scores

---

**Now load the new data into the panel using the 'Load Selected Data' button.**
Press the 'Project Data Into Model' to project the new data into the PCA model.
This brings up a new panel where you can plot the new scores alone or combined with the model scores.

The loadings are not shown in this panel since they are the same as the model loadings and can be explored using the 'Plot Loadings' function in the 'Data Display' menu.
Let's look at the scores of the new data first.
Data Selection Panel
These are the main input data that will be used in further analysis unless you specify otherwise.
Use the drop down menus to select the data and information you want to use in your analysis.

Name of Data Matrix: ntestdataset
Name of Variable Matrix: exactmass
Name of Filename Matrix: testfiles
Name of Totalcounts Matrix: totalcounts
Name of Samplenames Matrix: testsamples

PCA Modelling Panel
Choose a Model to Load
pcamodel

Select the data you want to project into the model, and corresponding samplenames from the 'Raw Data Selection Panel'

Load Selected Data

Add New Data: 
ntestdataset

Add New Samplenames: 
testsamples

Project Data into Model

Save Model Data to Workspace

Name for New Data Scores

Name for combined scores (new and model)

Save Combined Scores

Press the 'Plot Scores - New Data' button to plot only the scores for the new data.
Here we have plotted PC 1 vs PC2 scores for the new data.
Data Selection Panel
These are the main input data that will be used in further analysis unless you specify otherwise. Use the drop down menus to select the data and information you want to use in your analysis.

Name of Data Matrix: testdataset
Name of Variable Matrix: exactmass
Name of Filename Matrix: testfiles
Name of Totalcounts Matrix: totalcounts
Name of Samplenames Matrix: testsamples

PCA Modelling Panel
Choose a Model to Load: pcamodel

Model Information
Original Data Set
Data: testdataset, variables: exactmass, samples: samplenames

Model Data
Scores: pcascores, Loads: pcaloads, % Variance: pcavariance, Scaling Choice: Mean Center

Select the data you want to project into the model, and corresponding samplenames from the 'Raw Data Selection Panel'

Save Model Data to Workspace
Name for New Data Scores
Name for combined scores (new and model)

Load Selected Data
Project Data Into Model
Save New Scores
Save Combined Scores

Now let's look at all of the scores data together.
You can see that most new data points fall almost exactly where the model data points are. At some point I may change the colors of the new scores and model scores to make them easier to see in the combined plot.

For now you can compare them by switching back and forth between the two plots.
Data Selection Panel

These are the main input data that will be used in further analysis unless you specify otherwise. Use the drop-down menus to select the data and information you want to use in your analysis.

Name of Data Matrix: testdataset
Name of Variable Matrix: exactmass
Name of Filename Matrix: testfiles
Name of Totalcounts Matrix: totalcounts
Name of Samplenames Matrix: testsamples

PCA Modelling Panel

Choose a Model to Load:
pcamodel

Select the data you want to project into the model, and corresponding samplenames from the 'Raw Data Selection Panel'.

Load Selected Data

Save Model Data to Workspace

Model Information

Original Data Set

Data: testdataset
Variables: exactmass
Samples: samplenames

Model Data

Scores: pcascores
Loads: pcaploads
% Variance: pcapvariance
Scaling Choice: Mean

Just for reference. Here are the new sample scores alone.

Plot Scores. New Data
Plot Scores. New and Model Data

Close Panel
Data Selection Panel
These are the main input data that will be used in further analysis unless you specify otherwise. Use the drop-down menus to select the data and information you want to use in your analysis.

Name of Data Matrix: ntestdataset
Name of Variable Matrix: exactmass
Name of Filename Matrix: testfiles
Name of Totcounts Matrix: totalcounts
Name of Samenames Matrix: testsamples

PCA Modelling Panel
Choose a Model to Load
pcamodel

Model Information
Original Data Set
Data: ntestdataset
Variables: exactmass
Samples: samenames

Model Data
Scores: pcamodel
Loads: pcamodel
% Variance: pcamodel
Scaling Choice: Mean

Select the data you want to project into the model, and corresponding sample names from the 'Raw Data Selection Panel'

New Data: ntestdataset
New Samenames: testsamples

Save Model Data to Workspace
Name for New Data Scores
Save New Scores
Name for combined scores (new and model)
Save Combined Scores

And the combined scores again

Close Panel
### Data Selection Panel

These are the main input data that will be used in further analysis unless you specify otherwise. Use the drop-down menus to select the data and information you want to use in your analysis.

<table>
<thead>
<tr>
<th>Name of Data Matrix</th>
<th>Name of Variable Matrix</th>
<th>Name of Filenames Matrix</th>
<th>Name of Totalcounts Matrix</th>
<th>Name of Samplenames Matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>intestdataset</td>
<td>exactmass</td>
<td>filename</td>
<td>totalcounts</td>
<td>testsamples</td>
</tr>
</tbody>
</table>

### PCA Modelling Panel

Choose a Model to Load

- pcamodel

Select the data you want to project into the model, and corresponding samenames from the 'Raw Data Selection Panel'.

**To save the scores for the new data only, enter a name for the new scores**

**Save Model Data to Workspace**

- Name for New Data Scores
- Name for combined scores (new and model)

**Project Data into Model**

**Plot Scores: New Data**

**Plot Scores: New and Model Data**

**Principal Component 2 Scores (21%)**

- X-Axis
- Y-Axis
and press the 'Save New Scores' button.
Data Selection Panel

These are the main input data that will be used in further analysis unless you specify otherwise. Use the drop-down menus to select the data and information you want to use in your analysis.

Name of Data Matrix: testdataset
Name of Variable Matrix: exactmass
Name of Filename Matrix: testfiles
Name of Totalcounts Matrix: totalcounts
Name of Samplenames Matrix: testsamples

PCA Modelling Panel

Choose a Model to Load:
pcamodel

Select the data you want to project into the model, and corresponding samplenames from the 'Raw Data Selection Panel'

Model Information

Original Data Set
Data: ndataset
Variables: exactmass
Samples: samplenames

Model Data
Scores: pcасore
Loads: pcaloads
% Variance: pcavariance
Scaling Choice: Mean Center

Load Selected Data
New Data: testdataset

Save Model Data to Workspace

Name for New Data Scores: newscores
Name for combined scores (new and model): combinedscores

and press the 'Save Combined Scores' button.

Plot Scores: New Data
Plot Scores: New and Model Data

Close Panel
Data Selection Panel

These are the main input data that will be used in further analysis unless you specify otherwise. Use the drop-down menus to select the data and information you want to use in your analysis.

<table>
<thead>
<tr>
<th>Name of Data Matrix</th>
<th>Name of Variable Matrix</th>
<th>Name of Filename Matrix</th>
<th>Name of Totalcounts Matrix</th>
<th>Name of Samples Matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>ndataset</td>
<td>exactmass</td>
<td>filenames</td>
<td>totalcounts</td>
<td>samplenames</td>
</tr>
</tbody>
</table>

That's it for this tutorial.

Press the green button on the left to go back to the previous step. Press the button the right to go back to the beginning of the tutorial.