

UNIVERSITÀ
DEGLI STUDI
DI PADOVA

DEPARTMENT OF
INDUSTRIAL ENGINEERING 

Machine Learning Laboratory #1

Academic year 2025-2026

Prof. Pierantonio Facco

CAPE-Lab, Computer-Aided Process Engineering Laboratory

Email: pierantonio.facco@unipd.it

URL: <https://research.dii.unipd.it/capelab/>

Summary

- Software:
 - PLS_Toolbox[®] for Matlab[®]
- **Hands-on-pc laboratory**
 - Example #1: health and alcohol consumption

Multivariate statistics software

- Matlab[®] (by MathWorks)
 - Statistics toolbox
 - PLS_Toolbox[®] (by Eigenvector research Inc.)*
 - http://www.eigenvector.com/software/pls_toolbox.htm
- R
 - free
- Python
 - free
- Other software...
 - SPSS
 - Stata
 - SAS
 - JMP



J. Proc. Cont. Vol. 6, No. 6, pp. 329–348, 1996
Copyright © 1996 Published by Elsevier Science Ltd
Printed in Great Britain. All rights reserved
0959-1524/96 \$15.00 + 0.00

PII: S0950-1524(96)00009-1

Full Papers

The process chemometrics approach to process monitoring and fault detection

Barry M. Wise and Neal B. Gallagher

Eigenvector Research, 830 Wapato Lake Road, Manson, WA 98831, USA

Received in revised form 15 March 1996

Chemometrics, the application of mathematical and statistical methods to the analysis of chemical data, is finding ever widening applications in the chemical process environment. This article reviews the chemometrics approach to chemical process monitoring and fault detection. These approaches rely on the formation of a mathematical/statistical model that is based on historical process data. New process data can then be compared with models of normal operation in order to detect a change in the system. Typical modelling approaches rely on principal components analysis, partial least squares and a variety of other chemometric methods. Applications where the ordered nature of the data is taken into account explicitly are also beginning to see use. This article reviews the state-of-the-art of process chemometrics and current trends in research and applications. Copyright © 1996 Published by Elsevier Science Ltd

*Wise, B. M., and N. B. Gallagher (1996). The process chemometrics approach to process monitoring and fault detection. *J. Process Control* **6**, 329–348.

Matlab[®] codes

- Principal component analysis PCA* on \mathbf{X} = singular value decomposition on the correlation of \mathbf{X}

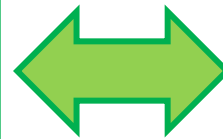
- $[U, S, V] = \text{svd}(\text{cov}(\text{auto}(\mathbf{X})))$;

- $\mathbf{P} = \mathbf{V}$; loadings \mathbf{P}

- $\mathbf{T} = \mathbf{X} * \mathbf{P}$; scores \mathbf{T}

$\left\{ \begin{array}{l} \text{autoscaled } \mathbf{X} \\ \mathbf{X} \text{ covariance} \end{array} \right. \quad \text{cov}(\mathbf{X}) = \frac{\mathbf{X}^T \mathbf{X}}{N - 1}$

$$\mathbf{X} = \mathbf{t}_1 \mathbf{p}_1^T + \mathbf{t}_2 \mathbf{p}_2^T + \dots + \mathbf{t}_k \mathbf{p}_k^T + \mathbf{E}$$



$$\text{cov}(\mathbf{X}) \mathbf{p}_i = \lambda_i \mathbf{p}_i$$

*Jackson, J. E. (2003). *A User's Guide to Principal Components*. Wiley-Interscience, Hoboken, N.J.

Geladi, P., and B.R., Kowalski 1986. Partial least-squares regression: a tutorial. *Anal. Chim. Acta* **185, 1–17.

PCA Matlab[®] code: NIPALS algorithm

```
function [T,P,E]=pca_profFacco(X)
T=[];
P=[];
for n=1:min(min(size(X)),20)
    n
    te=1;
    tn=1;
    if n==1
        X=X;
    else
        X=E;
    end
    m=1;
    while (m<=1000)&&((norm(te)/norm(tn))>10^-12)
        m
        if m==1000
            display('no convergence');
            norm(te)/norm(tn)
        else
            if m==1
                xn=X(:,1);
                tn=xn;
            else
                tn=t1;
            end
            p1=((tn'*X)/(tn'*tn))';
            p1=p1/norm(p1);
            t1=(X*p1)/(p1'*p1);
            te=t1-tn;
            m=m+1;
        end
    end
    T=horzcat(T,t1);
    P=horzcat(P,p1);
    E=X-t1*p1';
end
```

Example #1

Health and alcohol consumption

Health and alcohol consumption

- Available data:

- 10 observations (on the rows)
 - nations
- 5 variables (on the columns)
 - wine, beer and liquor consumption, heart disease rate life expectancy

	liquor consumption [L/year]	wine consumption [L/year]	beer consumption [L/year]	life expectancy [years]	heart disease rate [10 ⁵ cases/year]
France	2.5	63.5	40.1	78	61.1
Italy	0.9	58.0	25.1	78	94.1
Switzerland	1.7	46.0	65.0	78	106.4
Australia	1.2	15.7	102.1	78	173.0
Great Britain	1.5	12.2	100.0	77	199.7
USA	2.0	8.9	87.8	76	176.0
Russia	3.8	2.7	17.1	69	373.6
Czech Republic	1.0	1.7	140.0	73	283.7
Japan	2.1	1.0	55.0	79	34.7
Mexico	0.8	0.2	50.4	73	36.4

- Objective: **exploratory analysis** of the data

Loading data in Matlab

■ Command:

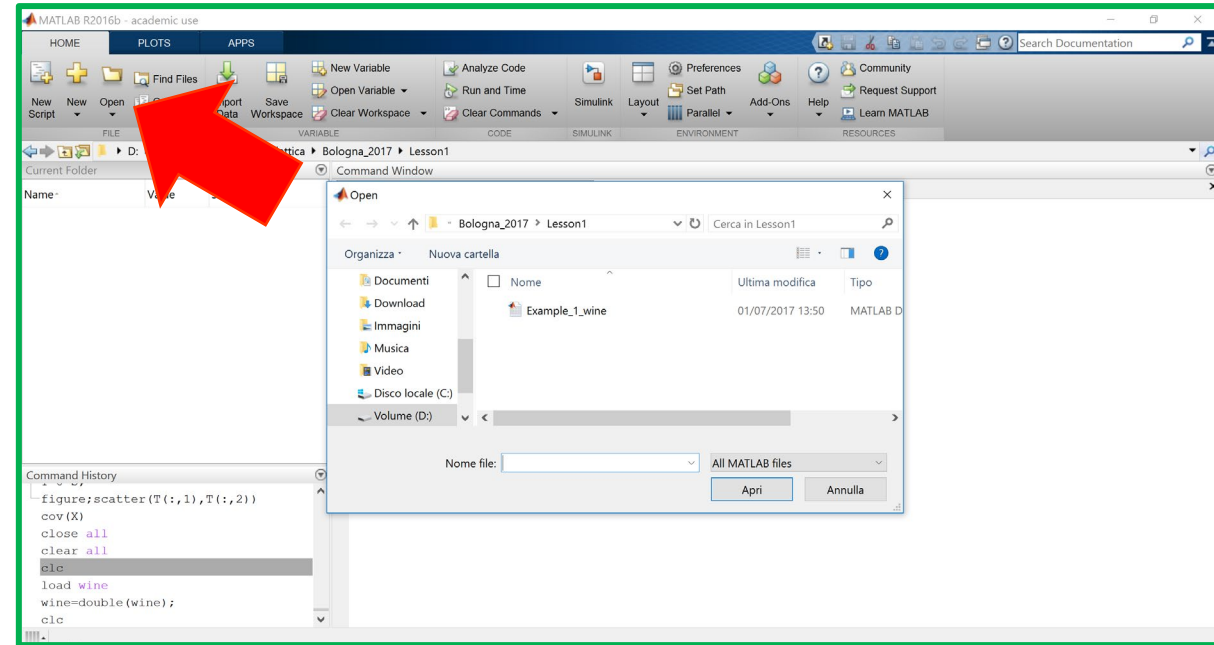
- `load wine`

*this is a **table**
dataset*

■ Alternative:

- Matlab® user interface
- select «Open»
- the path for the dataset `wine.mat`
 - shared in Moodle

*this is a **double**
precision dataset*



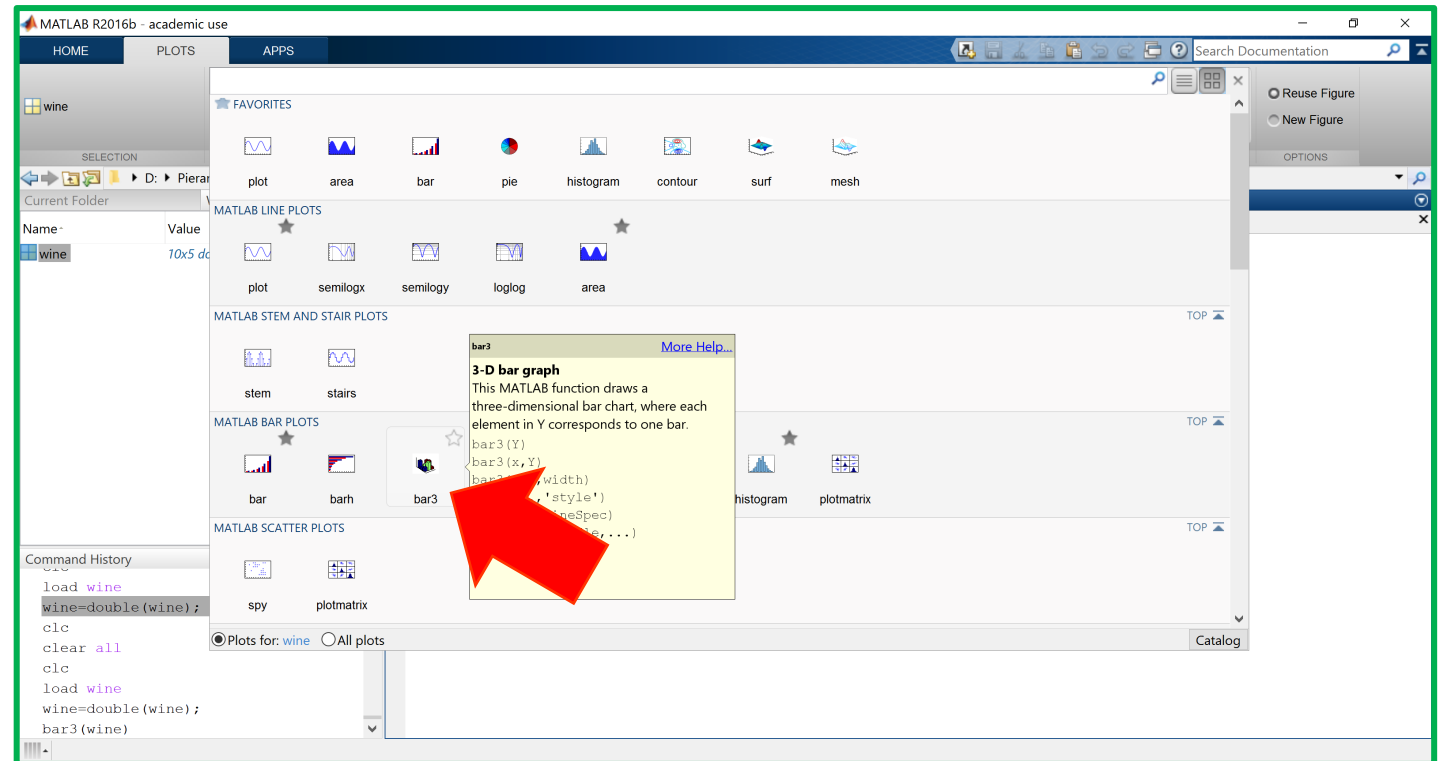
Initial data visualization

■ Commands:

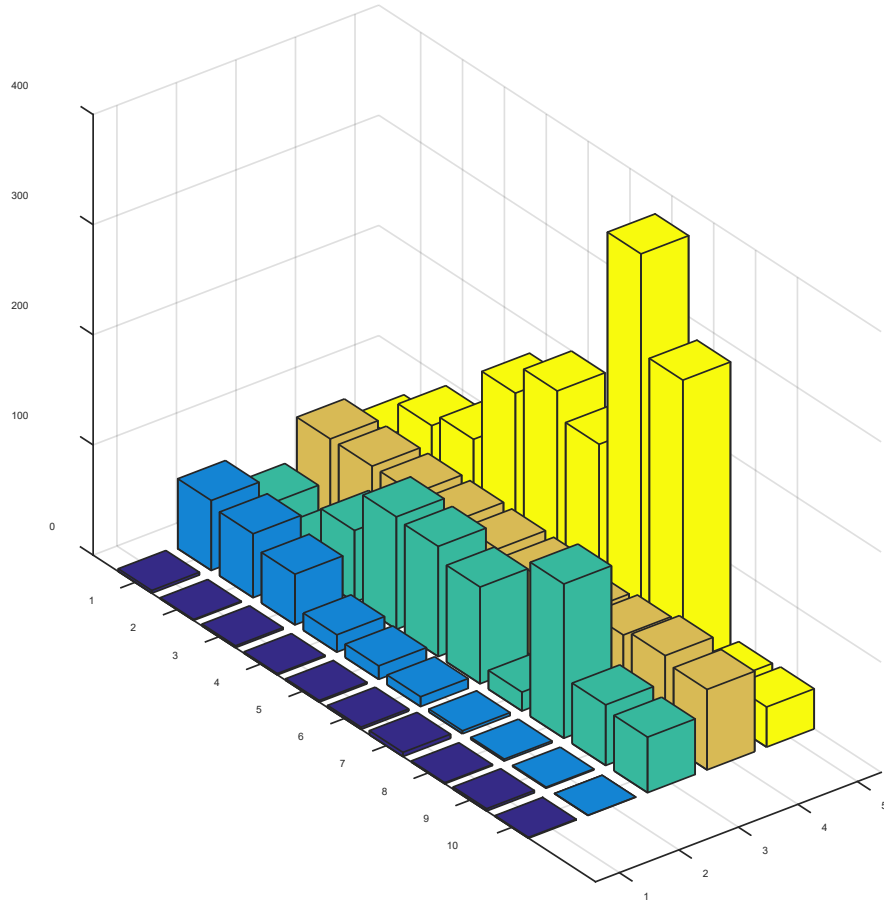
- `bar3(wine)` or `bar3(double(wine))`

■ User interface:

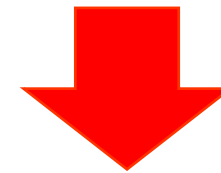
- select the dataset
- «Plot» → «bar3»



Visual data exploration



data have different scales



AUTOSCALING to give the same weight to all the data (same importance to all the variables)

PCA in the PLS_Toolbox

Command:

- `pca`

The screenshot displays the MATLAB R2016b interface. The Command Window shows the execution of the `pca` command, resulting in a variable `wine` of size `10x5 double`. The Analysis - PCA (No Model) (2) window is open, showing the Analysis Flowchart with steps: 1. Load calibration data, 2. Choose Preprocessing, 3. Choose Cross-Validation, 4. Build Model, 5. Choose Components, 6. Review Scores, 7. Review Loadings, 8. Load Test Data, and 9. Apply Model. The SSQ Table shows the Percent Variance Captured by PCA Model (* = suggested).

Number PCs	eigenvalue	% variance	% variance
1	1.4500	21.00	21.00

Select PCA

Analysis - PCA (No Model)

File Edit Preprocess Analysis Refine Tools Help FigBrowser

DECOMPOSITION

- ✓ PCA - Principal Component Analysis
 - Purity - SIMPLISMA
 - MCR - Multivariate Curve Resolution
 - MPCA - Multiway PCA
 - Batch Maturity
- CLUSTERING >
- REGRESSION >
- CLASSIFICATION >
- STATISTICAL >
- MULTI-WAY >
- DATA FUSION >
- Analysis Methods Help

View: []

Number PCs: []

Eigenvalue

	Eigenvalue	RMSECV
1		

RMSECV

Apply / Validate

Prediction

Analysis Flowchart

1. Load calibration data
2. Choose Preprocessing
3. Choose Cross-Validation
4. Build Model

Review Model

5. Choose Components
6. Review Scores
7. Review Loadings

Use Model

8. Load Test Data
9. Apply Model

Cache: "general" DATE View (* = Not Available)

- Cache Settings and View
- Demo Data
 - 01-Jul-2017
 - 28-Jun-2017
 - 20-Jun-2017
 - 15-Jun-2017
 - 05-Jun-2017
 - 18-May-2017
 - 16-May-2017
 - 27-Apr-2017
 - 19-Apr-2017
 - 07-Apr-2017
 - 04-Apr-2017
 - 03-Apr-2017

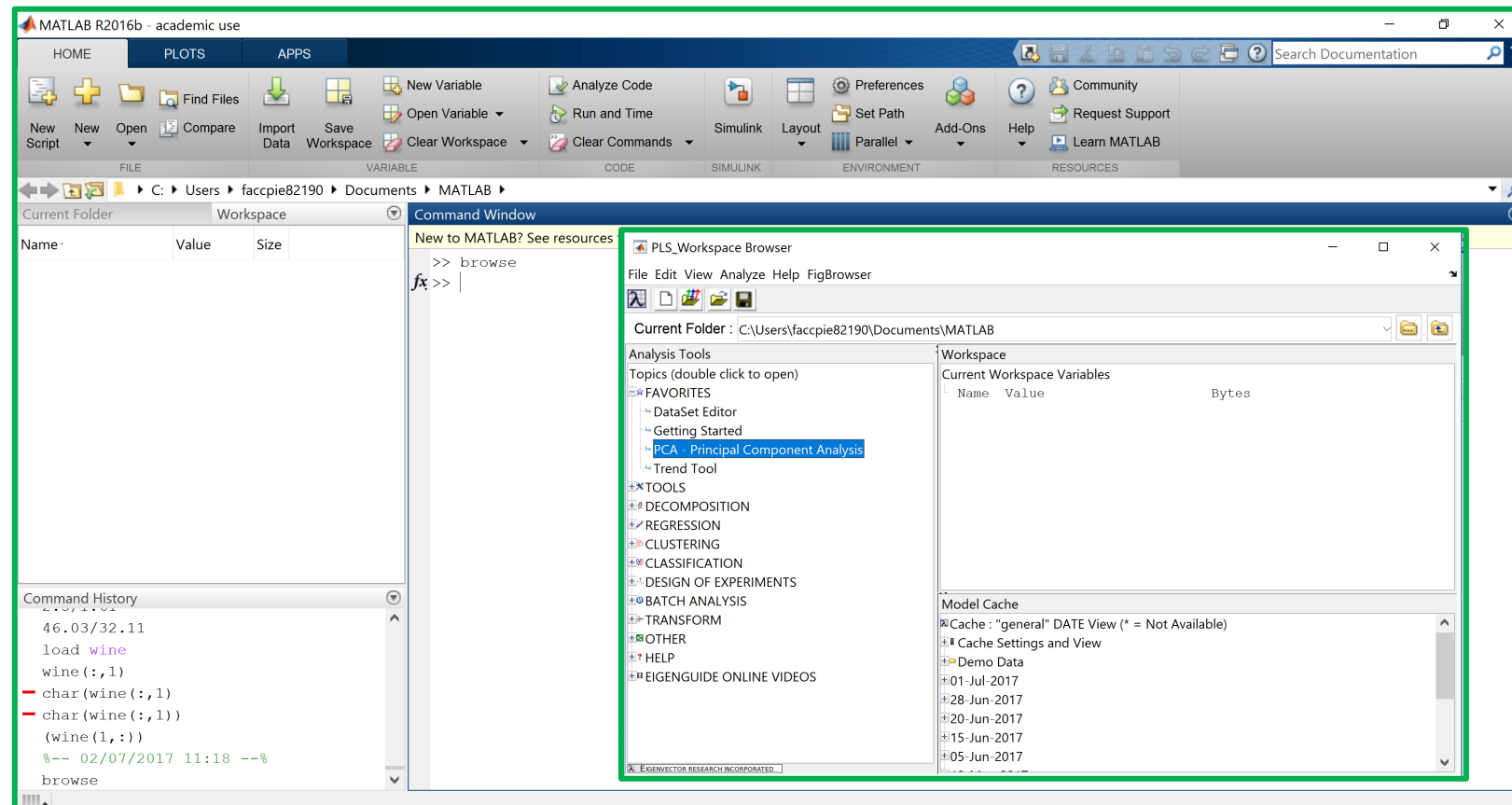
Analysis Help Pane: Neither data nor a model is currently loaded. Data can be loaded or imported and/or a model can be loaded using the File menu.

Alternative method

■ Command:

- **browse**

... and select PCA



Data upload in the toolbox

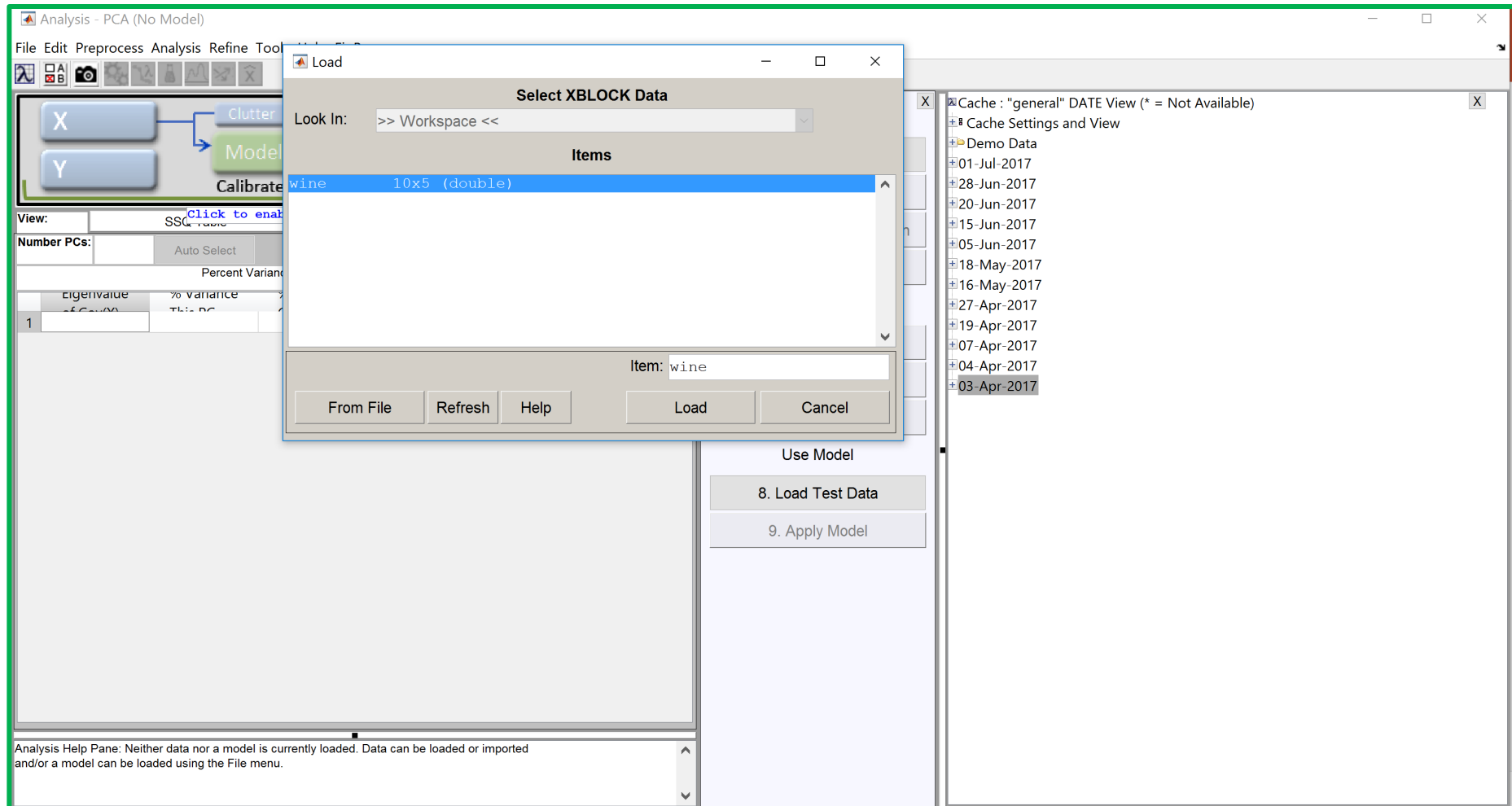
(1/2)

The screenshot displays the software interface for 'Analysis - PCA (No Model) (2)'. The 'File' menu is open, showing options like 'Load Data', 'Import Data', and 'Save Data'. The 'Load Data' sub-menu is expanded, highlighting 'X-Block'. In the background, a flowchart shows 'X' and 'Y' blocks leading to a 'Prediction' block, with an 'Apply / Validate' button. On the right, the 'Analysis Flowchart' panel lists steps: 1. Load calibration data, 2. Choose Preprocessing, 3. Choose Cross-Validation, 4. Build Model, 5. Choose Components, 6. Review Scores, 7. Review Loadings, 8. Load Test Data, and 9. Apply Model. The far right panel shows a 'Cache' with a list of dates from 2017.

Analysis Help Pane: Neither data nor a model is currently loaded. Data can be loaded or imported and/or a model can be loaded using the File menu.

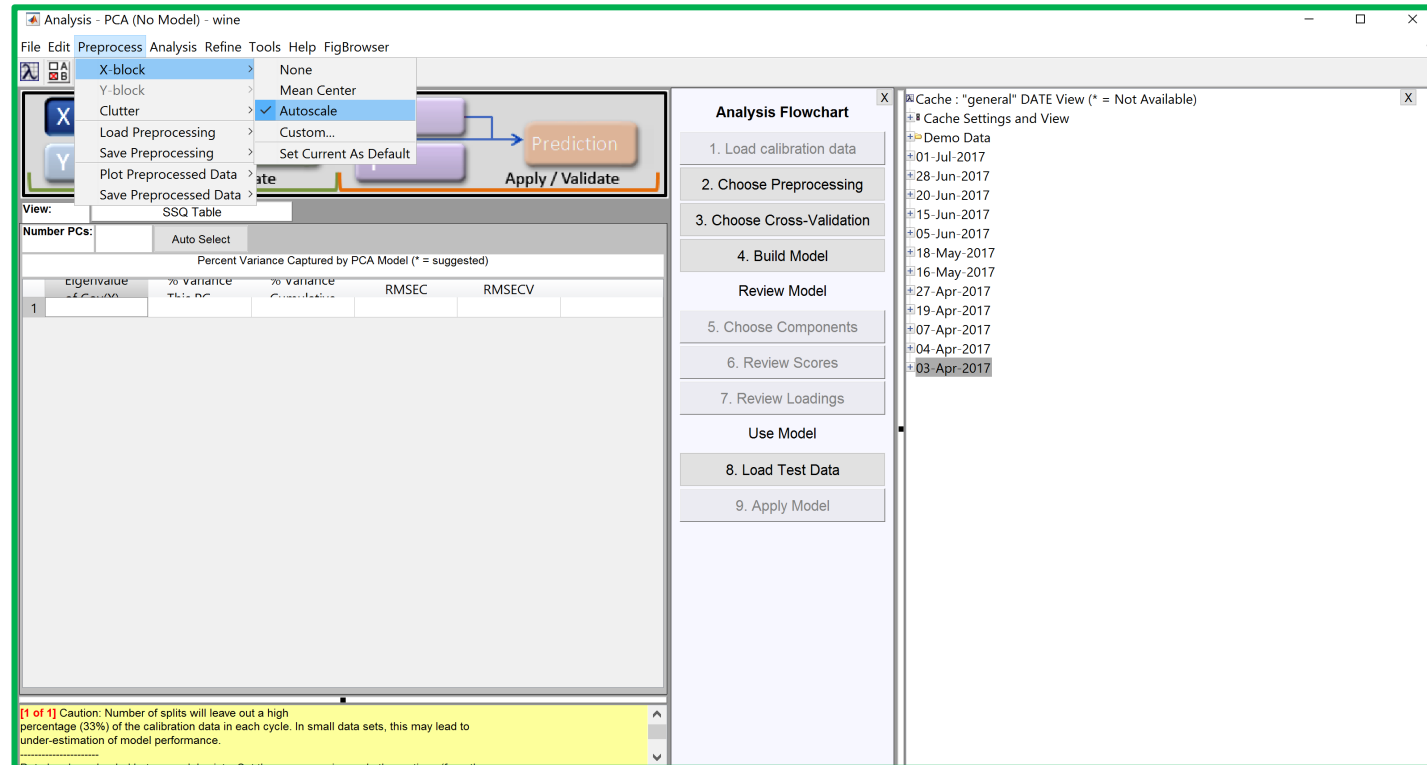
Data upload in the toolbox

(2/2)



Select pretreatment

- Variable autoscaling
 - mean centering
 - scaling to unit variance



Analysis - PCA (No Model) - wine

File Edit Preprocess Analysis Refine Tools Help FigBrowser

X-block None
Y-block Mean Center
Clutter Autoscale
Load Preprocessing Custom...
Save Preprocessing Set Current As Default
Plot Preprocessed Data
Save Preprocessed Data

View: SSQ Table

Number PCs: Auto Select

Percent Variance Captured by PCA Model (* = suggested)

Eigenvalue	% Variance	% Variance	RMSEC	RMSECV
1				

[1 of 1] Caution: Number of splits will leave out a high percentage (33%) of the calibration data in each cycle. In small data sets, this may lead to under-estimation of model performance.

Analysis Flowchart

1. Load calibration data
2. Choose Preprocessing
3. Choose Cross-Validation
4. Build Model

Review Model

5. Choose Components
6. Review Scores
7. Review Loadings

Use Model

8. Load Test Data
9. Apply Model

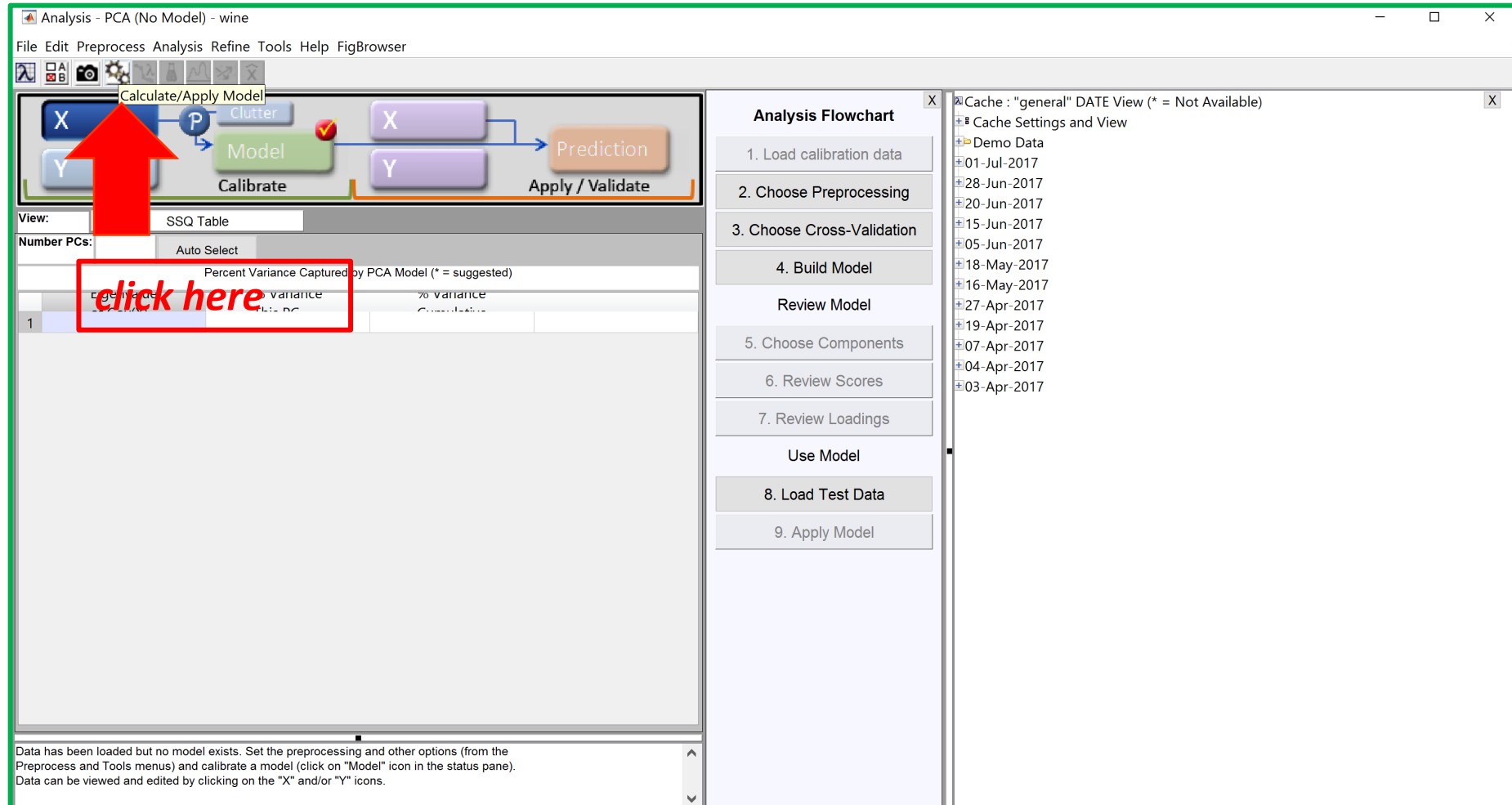
Cache: "general" DATE View (* = Not Available)

Cache Settings and View

Demo Data

- 01-Jul-2017
- 28-Jun-2017
- 20-Jun-2017
- 15-Jun-2017
- 05-Jun-2017
- 18-May-2017
- 16-May-2017
- 27-Apr-2017
- 19-Apr-2017
- 07-Apr-2017
- 04-Apr-2017
- 03-Apr-2017

PCA model building



Analysis - PCA (No Model) - wine

File Edit Preprocess Analysis Refine Tools Help FigBrowser

Calculate/Apply Model

X Clutter Y Model Prediction

Calibrate Apply / Validate

View: SSQ Table

Number PCs: Auto Select

Percent Variance Captured by PCA Model (* = suggested)

	variance	% variance
1		

click here

Analysis Flowchart

1. Load calibration data
2. Choose Preprocessing
3. Choose Cross-Validation
4. Build Model

Review Model

5. Choose Components
6. Review Scores
7. Review Loadings

Use Model

8. Load Test Data
9. Apply Model

Cache : "general" DATE View (* = Not Available)

- Cache Settings and View
- Demo Data
- 01-Jul-2017
- 28-Jun-2017
- 20-Jun-2017
- 15-Jun-2017
- 05-Jun-2017
- 18-May-2017
- 16-May-2017
- 27-Apr-2017
- 19-Apr-2017
- 07-Apr-2017
- 04-Apr-2017
- 03-Apr-2017

Data has been loaded but no model exists. Set the preprocessing and other options (from the Preprocess and Tools menus) and calibrate a model (click on "Model" icon in the status pane). Data can be viewed and edited by clicking on the "X" and/or "Y" icons.

Eigenvalues and explained variance

Analysis - PCA 1 PCs - wine

File Edit Preprocess Analysis Refine Tools Help FigBrowser

Analysis Flowchart

1. Load calibration data
2. Choose Preprocessing
3. Choose Validation
4. Build Model
5. Choose Components
6. Review Scores
7. Review Loadings
8. Load Test Data
9. Apply Model

View: SSQ Table

Number PCs: 1 Auto Select

Percent Variance Captured by PCA Model (* = suggested)

	eigenvalue	% variance	% variance	RMSEC	RMSECV	
	(λ^2)	This PC	Cumulative			current
1	2.30e+00	46.03	46.03	0.697	1.451	
2	1.61e+00	32.11	78.14	0.4435	1.755	
3	5.84e-01	11.68	89.83	0.3026	3.692	
4	4.22e-01	8.44	98.27	0.1247	21.65	
5	8.64e-02	1.73	100.00	3.028e-15	21.65	

A model has been calibrated from the data. Review the model using the toolbar button(s), save the model (File menu), or load test (validation) data (File menu). The number of components, preprocessing options, and other settings can also be modified to adjust the model. The data can be viewed and edited from the Edit menu.

PC	eigenvalue	explained variance (%)	cumulative explained variance (%)
1	2.30	46.03	46.03
2	1.61	32.11	78.14
3	0.58	11.68	89.82
4	0.42	8.44	98.27
5	0.09	1.73	100.00

Select the right number of PCs

Analysis - PCA 1 PCs - wine

File Edit Preprocess Analysis Refine Tools Help FigBrowser

Plot Eigenvalues/Cross-validation results

Model Calibrate Apply / Validate Prediction

Number PCs: 1

	Eigenvalue	% variance	% variance	RMSECV	RMSECV	
1	2.30e+00	46.03	46.03	0.697	1.451	current
2	1.61e+00	32.11	78.14	0.4435	1.755	
3	5.84e-01	11.68	89.83	0.3026	3.692	
4	4.22e-01	8.44	98.27	0.1247	21.65	
5	8.64e-02	1.73	100.00	3.028e-15	21.65	

model has been calibrated from the data. Review the model using the toolbar button(s), save the model (File menu), or load test (validation) data (File menu). The number of components, preprocessing options, and other settings can also be modified to adjust the model. The data can be viewed and edited from the Edit menu.

Analysis Flowchart

1. Load calibration data
2. Choose Preprocessing
3. Choose Cross-Validation
4. Build Model

Review Model

5. Choose Components
6. Review Scores
7. Review Loadings

Use Model

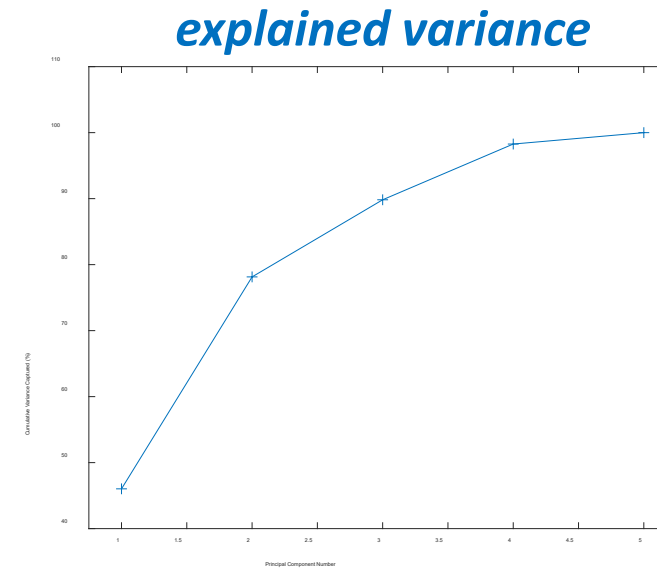
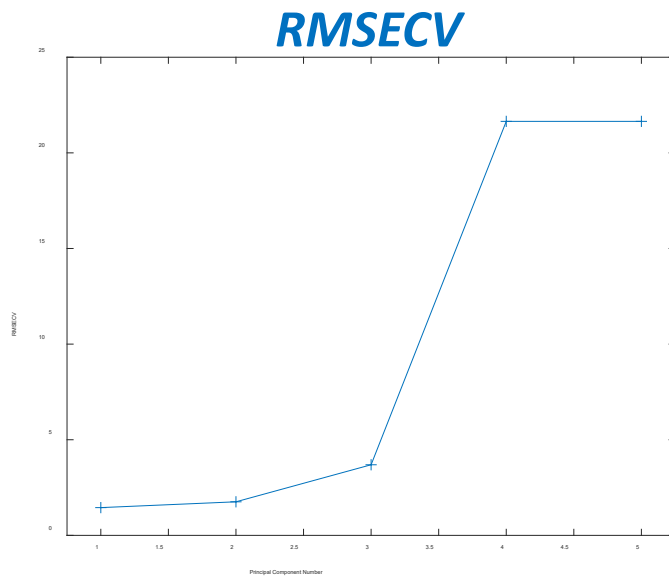
8. Load Test Data
9. Apply Model

Cache : "general" DATE View (* = Not Available)

- Cache Settings and View
- Demo Data
- 01-Jul-2017
- 28-Jun-2017
- 20-Jun-2017
- 15-Jun-2017
- 05-Jun-2017
- 18-May-2017
- 16-May-2017
- 27-Apr-2017
- 19-Apr-2017
- 07-Apr-2017
- 04-Apr-2017
- 03-Apr-2017

How to select the correct number of PCs

- Eigenvalue >1
- Scree test
 - identify the “elbow” on the figure of the explained variance vs. number of PCs
- Cross validation
 - identify the minimum of the RMSECV



Valle, S., W. Li, W., S. J. Qin (1999). Selection of the Number of Principal Components: The Variance of the Reconstruction Error Criterion with a Comparison to Other Methods. *Ind. Eng. Chem. Res.* **38**, 4389-4401

Loading plot

Analysis - PCA 2 PCs - wine

File Edit Preprocess Analysis Refine Tools Help FigBrowser

Plot loads and variable statistics

Calibrate Apply / Validate

View: Selection

Number PCs: 2

	eigenvalue	% variance	% variance	RMSEC	RMSECV	
1	2.30e+00	46.03	46.03	0.697	1.451	
2	1.61e+00	32.11	78.14	0.4435	1.755	current
3	8.4e-01	11.68	89.83	0.3026	3.692	
4	4.01e-01	8.44	98.27	0.1247	21.65	
5	1.73e-01	1.73	100.00	3.028e-15	21.65	

Analysis Flowchart

1. Load calibration data
2. Choose Preprocessing
3. Choose Cross-Validation
4. Build Model
5. Choose Components
6. Review Scores
7. Review Loadings
8. Load Test Data
9. Apply Model

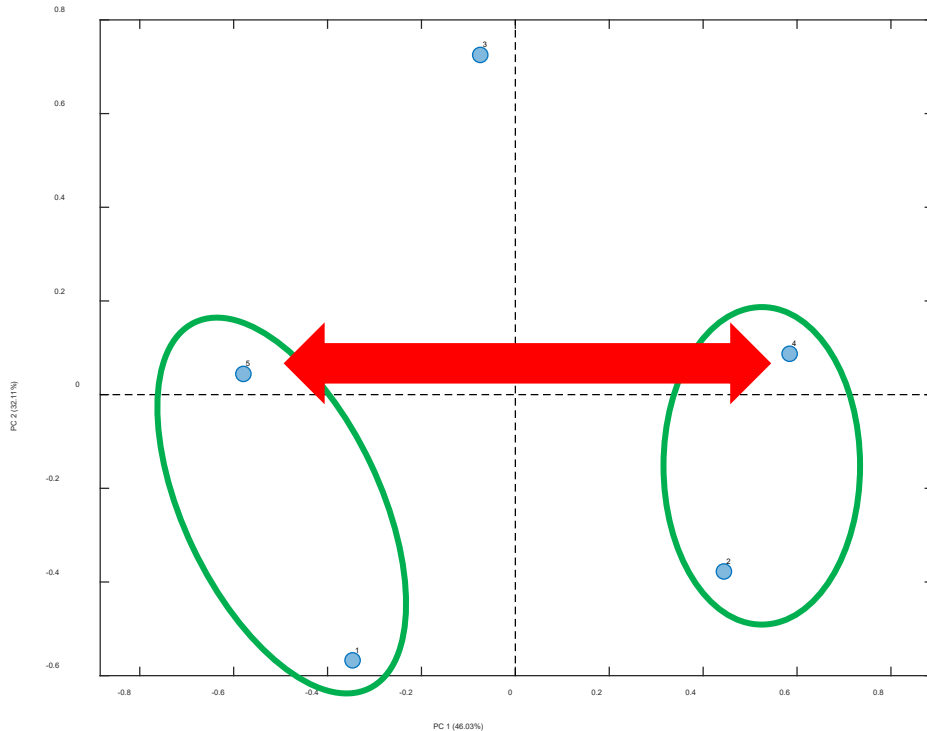
Cache : "general" DATE View (* = Not Available)

Demo Data

- 01-Jul-2017
- 28-Jun-2017
- 20-Jun-2017
- 15-Jun-2017
- 05-Jun-2017
- 18-May-2017
- 16-May-2017
- 27-Apr-2017
- 19-Apr-2017
- 07-Apr-2017
- 04-Apr-2017
- 03-Apr-2017

A model has been calibrated from the data. Review the model using the toolbar button(s), save the model (File menu), or load test (validation) data (File menu). The number of components, preprocessing options, and other settings can also be modified to adjust the model. The data can be viewed and edited from the Edit menu.

Variables correlation



	variable	units
1	liquor consumption	L/year
2	wine consumption	L/year
3	beer consumption	L/year
4	life expectation	years
5	heart disease frequency	1/(100000 year)

■ PC1 (46% explained variance)

• positive correlation

- variables 2-4: wine consumption – life expectation
- variables 1-5: liquor consumption – heart disease frequency

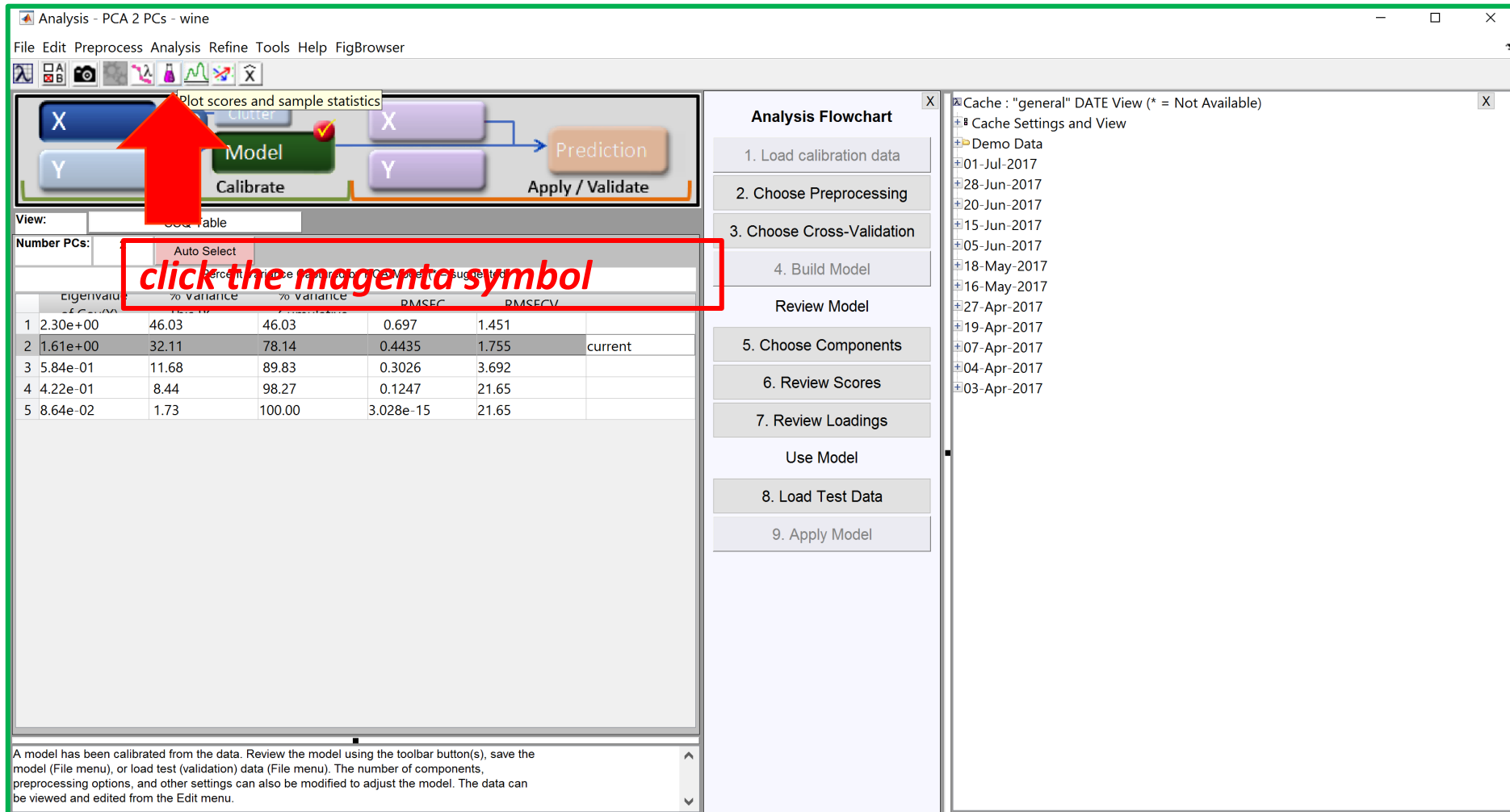
• anti-correlation

- 2-4/1-5: wine consumption(+ life expectation) and heart disease frequency (and liquor consumption)

■ PC2 (32%)

- beer consumption seems to be independent life expectation and heart disease frequency and partially anti-correlated to wine and liquor consumption

Score plot



Analysis - PCA 2 PCs - wine

File Edit Preprocess Analysis Refine Tools Help FigBrowser

Plot scores and sample statistics

Calibrate Model Apply / Validate Prediction

View: [table]

Number PCs:	Auto Select	Percent Variance Explained	Suggested			
		% variance	% variance	RMSEC	RMSECV	
1	2.30e+00	46.03	46.03	0.697	1.451	
2	1.61e+00	32.11	78.14	0.4435	1.755	current
3	5.84e-01	11.68	89.83	0.3026	3.692	
4	4.22e-01	8.44	98.27	0.1247	21.65	
5	8.64e-02	1.73	100.00	3.028e-15	21.65	

Analysis Flowchart

1. Load calibration data
2. Choose Preprocessing
3. Choose Cross-Validation
4. Build Model
- Review Model
5. Choose Components
6. Review Scores
7. Review Loadings
- Use Model
8. Load Test Data
9. Apply Model

Cache : "general" DATE View (* = Not Available)

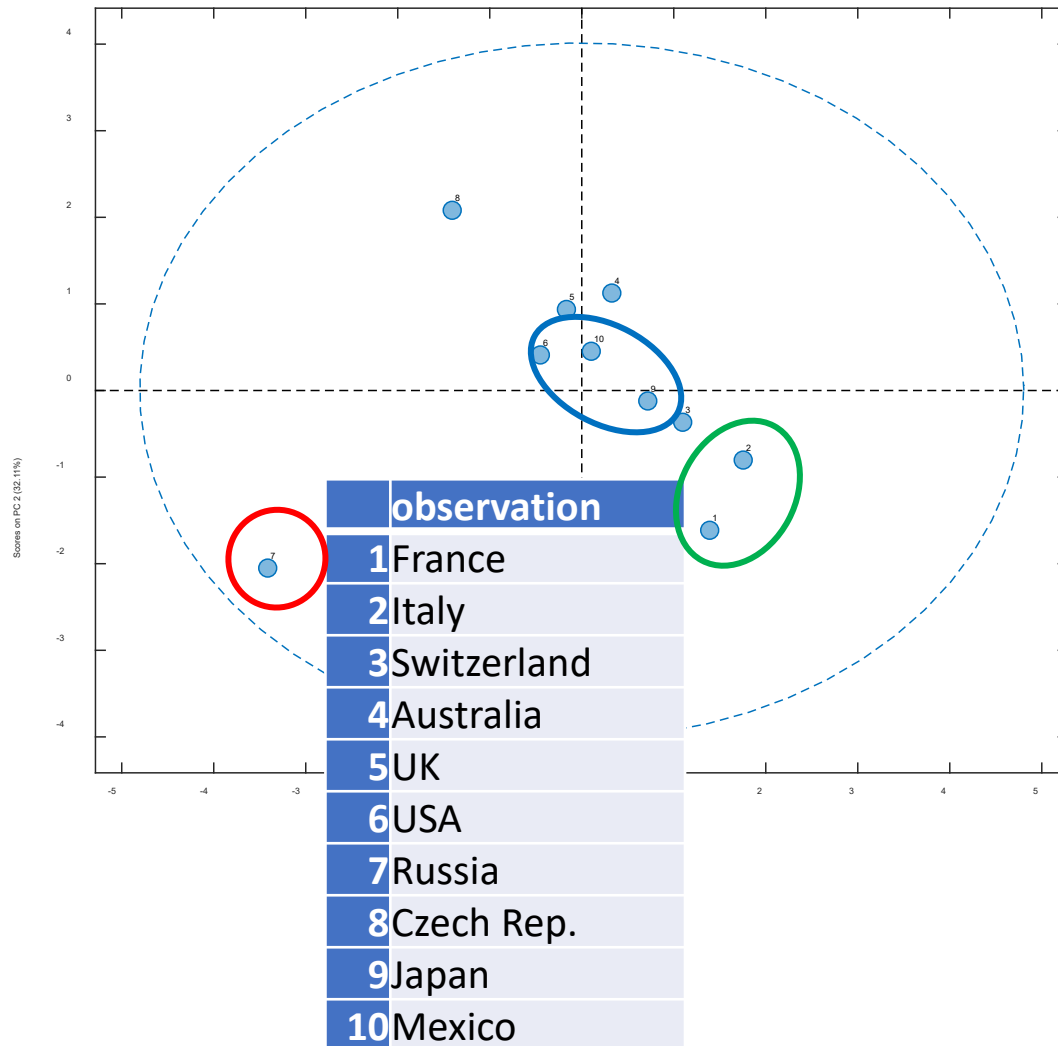
Cache Settings and View

Demo Data

- 01-Jul-2017
- 28-Jun-2017
- 20-Jun-2017
- 15-Jun-2017
- 05-Jun-2017
- 18-May-2017
- 16-May-2017
- 27-Apr-2017
- 19-Apr-2017
- 07-Apr-2017
- 04-Apr-2017
- 03-Apr-2017

A model has been calibrated from the data. Review the model using the toolbar button(s), save the model (File menu), or load test (validation) data (File menu). The number of components, preprocessing options, and other settings can also be modified to adjust the model. The data can be viewed and edited from the Edit menu.

Relations among observations



- France (1) and Italy (2), are **very similar**: high PC1
- Russia (7) is **very dissimilar** from other nations on both PC1 and PC2
- Czech Rep. (8) display the highest PC2
- USA (6), Japan and Mexico (10) are **“average observations”** because they are close to the origin

Biplot

Analysis - PCA 2 PCs - Wine

File Edit Preprocess Analysis Refine Tools Help FigBrowser

Workflow diagram: X, Y, Model, Apply / Validate, Prediction

View: **PCA Table**

Number PCs: 2

	Eigenvalue of Cov(M)	% variance This PC	% variance Cumulative	RMSEC	RMSECV	
1	2.30e+00	46.03	46.03	0.697	1.49	
2	1.61e+00	32.11	78.14	0.4435	1.898	current
3	5.84e-01	11.68	89.83	0.3026	4.121	
4	4.22e-01	8.44	98.27	0.1247	6.173	
5	8.64e-02	1.73	100.00	3.028e-15	6.173	

Analysis Flowchart

1. Load calibration data
2. Choose Preprocessing
3. Choose Cross-Validation
4. Build Model

Review Model

5. Choose Components
6. Review Scores
7. Review Loadings

Use Model

8. Load Test Data
9. Apply Model

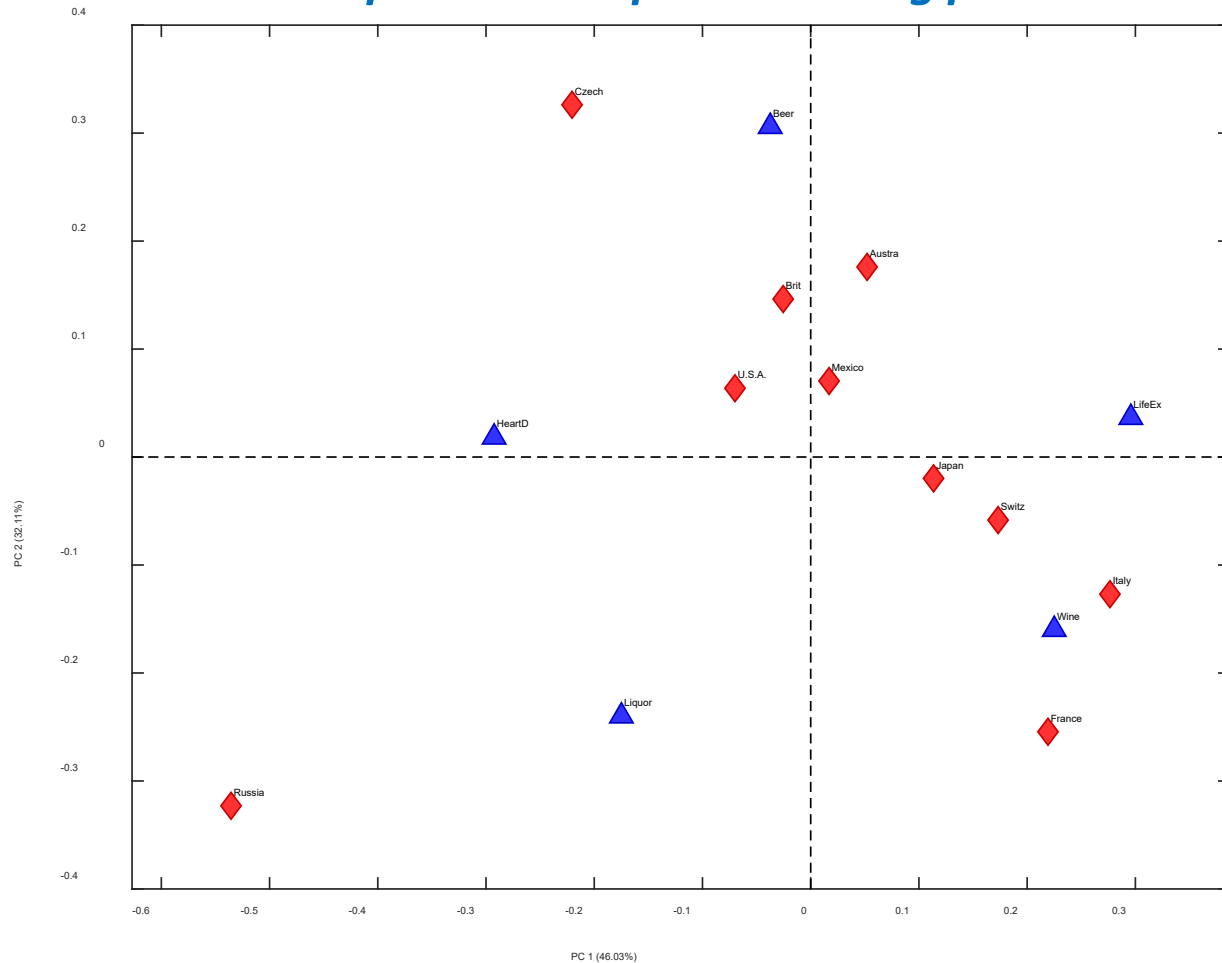
Cache: "general" DATE View (* = Not Available)

- Cache Settings and View
- Demo Data
- 01-Jul-2017
- 28-Jun-2017
- 20-Jun-2017
- 15-Jun-2017
- 05-Jun-2017
- 18-May-2017
- 16-May-2017
- 27-Apr-2017
- 19-Apr-2017
- 07-Apr-2017
- 04-Apr-2017
- 03-Apr-2017
- item: PCA (svd) 4 PCs [Autoscale] 2017-04-03 16:22:03.56
- item: X4 [42,21]
- item: PCA (svd) 4 PCs [Autoscale] 2017-04-03 16:20:09.11
- item: X3 [82,19]
- item: PCA (svd) 4 PCs [Autoscale] 2017-04-03 16:17:49.46
- item: X2 [193,20]
- item: X2 [193,20]
- item: PCA (svd) 4 PCs [Autoscale] 2017-04-03 16:15:21.40
- item: X1 [277,13]
- item: PCA (svd) 3 PCs [Autoscale] 2017-04-03 16:15:16.83
- item: PCA (svd) 1 PCs [Autoscale] 2017-04-03 16:14:46.39
- item: X1 [277,13]
- item: PCA (svd) 3 PCs [Autoscale] 2017-04-03 13:04:11.94
- item: X1 [277,20]
- item: PCA (svd) 1 PCs [Autoscale] 2017-04-03 13:04:05.72
- item: X1 [277,20]

[1 of 1] Caution: Number of splits will leave out a high percentage (33%) of the calibration data in each cycle. In small data sets, this may lead to under-estimation of model performance.

Biplot: joint score and loading reading

bi-plot = score plot + loading plot



- Japan, Italy, Switzerland and France are the nations with higher life expectancy
- France and Italy have the highest wine consumption
- Russia has the lowest life expectancy and the highest rate of heart disease and liquor consumption (but low wine and beer consumption)
- Czech Rep. has the highest beer consumption

	liquor consumption	wine consumption	beer consumption	life expectancy	heart disease rate
France	2.5	63.5	40.1	78	61.1
Italy	0.9	58.0	25.1	78	94.1
Switzerland	1.7	46.0	65.0	78	106.4
Australia	1.2	15.7	102.1	78	173.0
Grait Britain	1.5	12.2	100.0	77	199.7
USA	2.0	8.9	87.8	76	176.0
Russia	3.8	2.7	17.1	69	373.6
Czech Republic	1.0	1.7	140.0	73	283.7
Japan	2.1	1.0	55.0	79	34.7
Mexico	0.8	0.2	50.4	73	36.4

Save PCA model

Analysis - PCA 2 PCs - Wine

File Edit Preprocess Analysis Refine Tools Help FigBrowser

Load Data >

Import Data >

New Data >

Load Model

Import Model

Load Prediction

Load Options

Save Data > SSQ Table

Save Model

Export Model >

Export Predictions >

Save Prediction

Save Options

Clear >

Close

Auto Select

Percent Variance Captured by PCA Model (* = suggested)

% variance	% variance	RMSEC		
03	46.03	0.697	1	
11	78.14	0.4435	1	
68	89.83	0.3026	3	
44	98.27	0.1247	2	
5 8.64e-02	1.73	100.00	3.028e-15	2

Look In: >> Workspace <<

Items

wine 10x5 (dataset)

Item: PCAmode11

To File Refresh Help Save Cancel

1. name the model

2. save

A model has been calibrated from the data. Review the model using the toolbar button(s), save the model (File menu), or load test (validation) data (File menu). The number of components, preprocessing options, and other settings can also be modified to adjust the model. The data can be viewed and edited from the Edit menu.

PCA model is saved in a data structure

The image shows the MATLAB R2016b interface. The workspace contains a PCA model structure named 'PCAmodel1' and the 'wine' dataset. The Command Window shows the execution of 'load wine' and 'pca'. An 'Open Model' dialog box is open, asking to load the model into the model-building interface or view its raw contents. The 'View Raw Contents' button is highlighted with a red arrow.

1. PCA model structure in the workspace (double-click it)

2. select "View raw content"

Access details on PCA model

The image shows the MATLAB R2016b interface with a PCA model object selected in the workspace. The 'Variables - model_content' window is open, displaying the fields of the model. A red arrow points to the 'detail' field, which is annotated with '4. click detail'. Three other red arrows point to the 'loads', 'tsqs', and 'ssqresiduals' fields, which are annotated with '1. loads = scores + loadings', '2. tsqs = Hotelling T²', and '3. ssqresiduals = residuals' respectively. The Command History window shows the commands used to create the PCA model.

Field	Value
modeltype	'PCA'
author	'faccpie82190@MWFACCPPIE-01'
datasource	1x1 cell
date	'22-Mar-2019'
time	[2019,3,22,12,0,48.2430]
info	'Scores are in cell 1 of the loads field.'
userdata	1x1 cell
detail	1x1 struct
loads	2x1 cell
pred	1x1 cell
tsqs	2x1 cell
ssqresiduals	2x1 cell
description	3x1 cell
help	1x1 struct
modelversion	'8.2'

```
Command History
scatter(1:10,E(:,3))
scatter(1:10,E(:,1))
mean(E(:,1))
scatter(1:10,E(:,1))
box on;xlabel('observation');yla...
%-- 22/03/2019 08:26 --%
a=randn(100,1);
b=randn(100);
-a*b*a'
a'*b*a
%-- 22/03/2019 11:59 --%
load wine
pca
```

```
Command Window
>> load wine
>> pca
fx >>
```

Other details about PCA

The image shows the MATLAB R2016b interface with the 'wine' dataset loaded. The 'model_content.detail' window displays the following fields and values:

Field	Value
history	75x1 cell
data	10x5 double
res	1x1 cell
ssq	5x1 double
eig	[1]
rmsec	[0.6970, 0.4435, 0.3026, 0.1247, 3.0277e-15]
rmsecv	[1.4511, 1.7551, 3.6918, 21.6468, 21.6468]
esterror	0x0 struct
means	1x1 cell
stds	1x1 cell
reslim	1x1 cell
tsqlim	1x1 cell
reseig	[0.1842, 0.4222, 0.0864]
eigsnr	[1]
cv	'vet'
split	5
iter	1
cvi	[1,2,3,4,5,1,2,3,4,5]

Red boxes and arrows highlight the following fields and their meanings:

- ssq** = eigenvalues + explained variance
- reslim** = residuals limit
- tsqlim** = Hotelling T^2 limit

The Command Window shows the following commands:

```
>> load wine
>> pca
fx >>
```

Model details and flexibility

- The model is available in the Matlab[®] *Workspace* in a *structure* with the assigned name
- A lot of useful information are stored in the structure:
 - PC explained variance
 - scores and loadings
 - model and sample diagnostics
 - confidence limits
 - etc...
- The model building and processing is **much faster writing and running Matlab[®] code**

PCA with PLS_Toolbox[®] and Matlab[®] commands

```
clear all
close all
clc

% load alcohol consumption dataset
load wine

% data visualization
X=double(wine);
bar3(wine);xlabel('variable');ylabel('country');

% PCA with the PLS Toolbox
opca.display='off';
opca.plots='none';
opca.preprocessing='autoscale';
pca1=pca(X,2,opca);

% scores
T=pca1.loads{1,1};
figure;scatter(T(:,1),T(:,2));box on;xlabel('PC1 score');ylabel('PC2 score')

% loadings
P=pca1.loads{2,1};
```

```
% PCA matrix reconstruction
Xhat=T*P';
Xa=auto(X);

% residuals
E=Xa-Xhat;

% eigenvalues and explained variance
lambda=pca1.detail.ssq(:,2);
2.3014/sum(lambda)
lambda./sum(lambda)
lambda./sum(lambda)*100
cumsum(lambda./sum(lambda)*100)

% Q residuals
Q=sum(E.^2,2);

% Hotelling T^2
T2=diag(T*inv(diag(lambda(1:2)))*T');

% singular value decomposition
[U,Lsvd,Psvd]=svd(cov(auto(X)));
Tsvd=Xa*Psvd;
```

Today homework

- Replicate through commands what we did with the PLS_Toolbox graphic user interface
 - `[PCAmode11]=pca (X, 2) ;`
 - `[predict]=pca (Xnew, PCAmode11) ;`
- Practice with command
 - `svd`
- Extract all the mathematical entities:
 - scores
 - loadings
 - residuals
 - eigenvalues
 - etc...

... per sempre a fianco a me!

