

## Partial Least-Squares Regression (PLSR) using SIMCA-P software

### **Format experimental data for analysis in SIMCA-P**

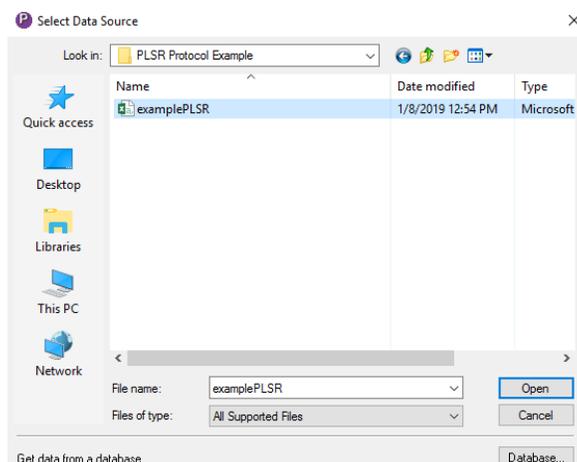
Mean-center and variance-scale your experimental data (example below). Save just the column and row labels, with normalized values, in a separate comma-separated value (.csv) file.

	A	B	C	D	E
1		X			Y
2		ERK	AKT	EGFR	death
3	DMSO	10	98	38	10
4	DRUG1	12	69	100	48
5	DRUG2	29	55	18	50
6	DRUG3	18	32	83	42
7					
8	mean	17.25	63.5	59.75	37.5
9	stdev	8.53912564	27.5983091	38.1957677	18.6458217
10					
11		X			Y
12		ERK	AKT	EGFR	death
13	DMSO	-0.8490331	1.25007658	-0.5694348	-1.4748613
14	DRUG1	-0.614817	0.19928757	1.05378167	0.56312884
15	DRUG2	1.3760191	-0.3079899	-1.093053	0.67039148
16	DRUG3	0.08783101	-1.1413743	0.60870618	0.24134093
17					

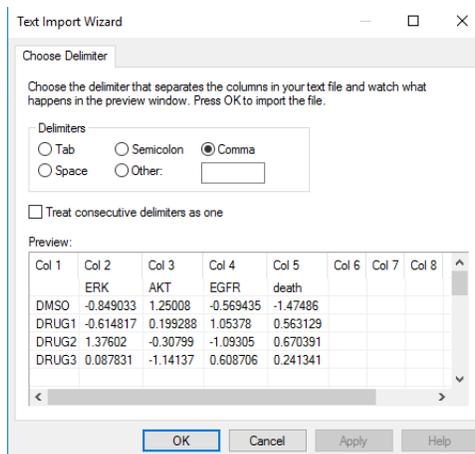
	ERK	AKT	EGFR	death
DMSO	-0.8490331	1.25007658	-0.5694348	-1.4748613
DRUG1	-0.614817	0.19928757	1.05378167	0.56312884
DRUG2	1.3760191	-0.3079899	-1.093053	0.67039148
DRUG3	0.08783101	-1.1413743	0.60870618	0.24134093

### **Open SIMCA-P and load experimental data (.csv file)**

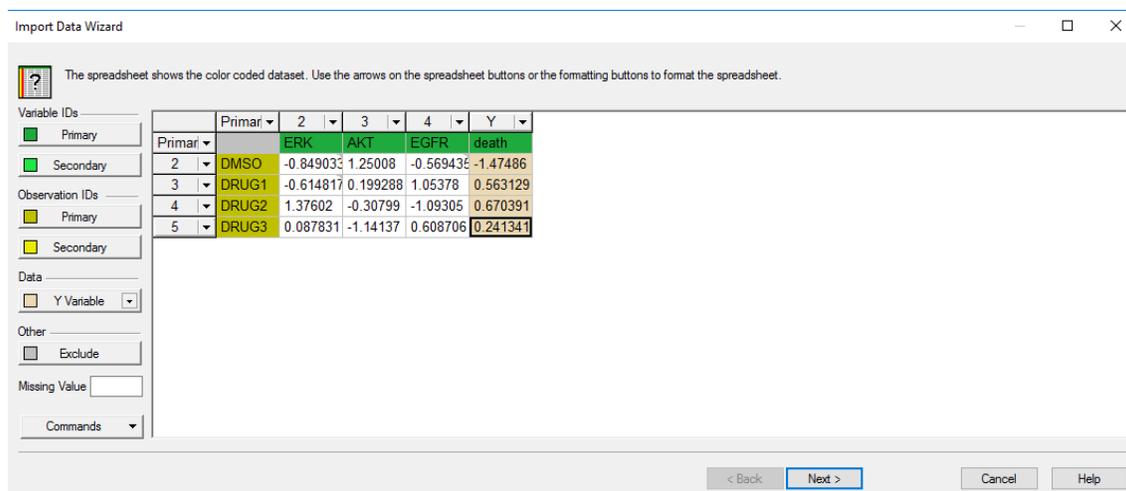
1. Open software (shortcut link on Lazzara Lab computer desktop) and generate a new project (File → New; ctrl+N).
2. In the “Select Data Source” dialog box, navigate to the .csv file of interest. Click “Open”.



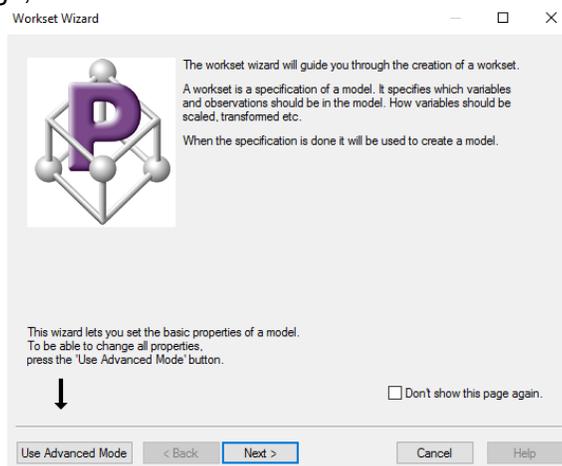
3. In the “Text Import Wizard” dialog box, select ‘Comma’ as the Delimiter option (or another if you decide to use a different data input file type). Make sure the data in the Preview window retains the format of your original .csv file. Click “OK”.



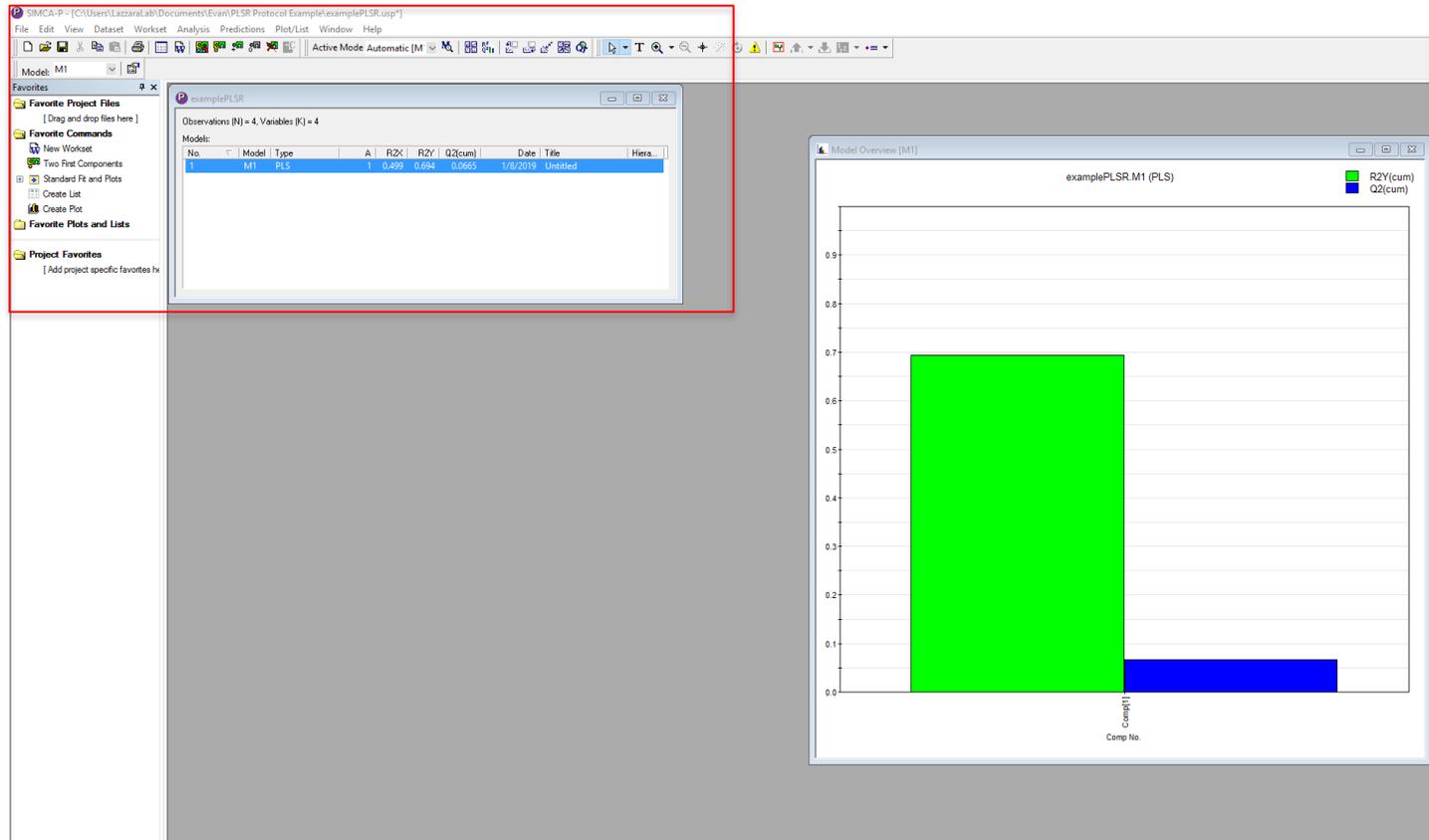
- In the "Import Data Wizard", specify which columns are y-variable (phenotypes) by highlighting the column and selecting the "Y Variable" option under the Data pull-down menu. The final table should look as below:



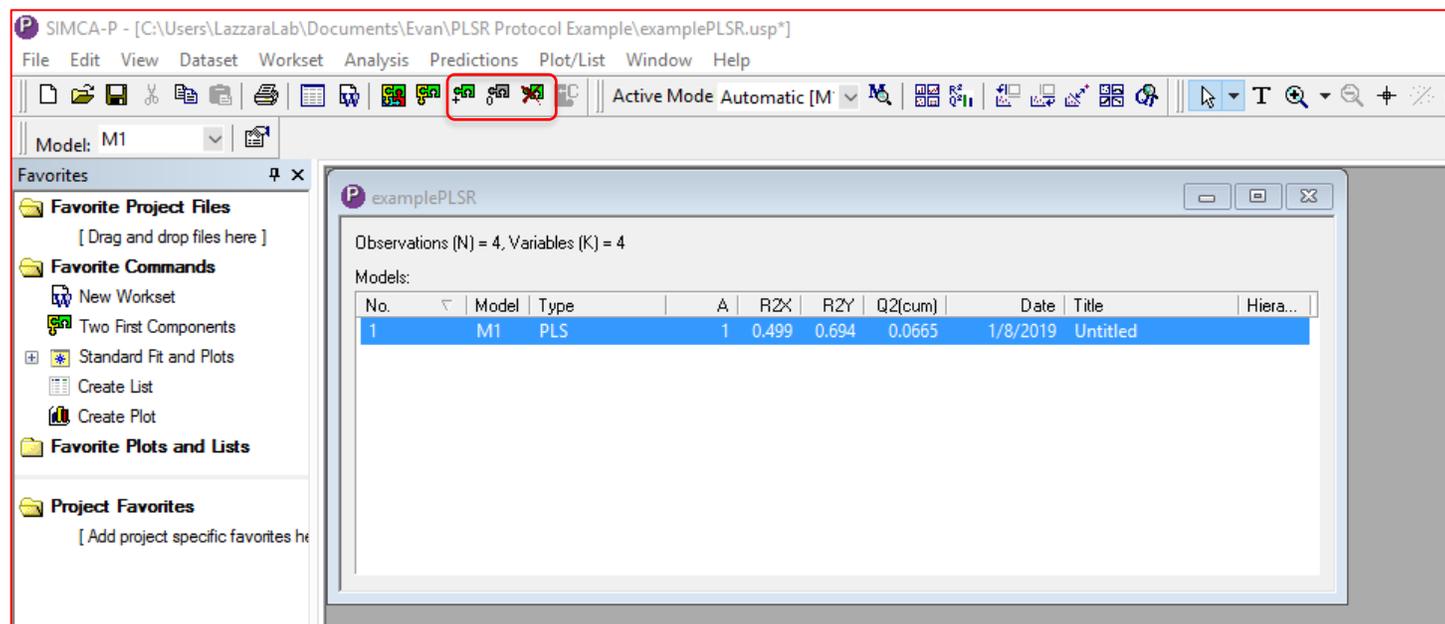
- Proceed through the remainder of the "Import Data Wizard" by clicking "Next" until the "Finish" button is available for selection. Use all default settings, unless desired otherwise. Click the "Finish" button.
- Next, advance through the "Workset Wizard" by clicking "Next" until the "Finish" button is available for selection. Use all default settings, unless desired otherwise.



7. After completing the “Workset Wizard” a Model Overview plot will appear (as shown below). The baseline model will only contain one principal component.



8. Add/subtract principal components to/from the model using the puzzle piece buttons. New columns will appear/disappear in the Model Overview plot as you add/subtract components.



- Double-click the row for your model under the box labeled after your .csv file (in this case “examplePLSR”) to bring up the full table of model statistics (R2X, R2Y, Q2, etc.) for each component.

The screenshot shows two windows from the SIMCA-P software. The top window, titled 'examplePLSR', displays a table of models. A red arrow points to the first row, which is highlighted in blue. A text box with the text 'Double click row' is positioned over the arrow. The bottom window, titled 'examplePLSR - M1', shows a detailed table of component statistics.

No.	Model	Type	A	R2X	R2Y	Q2(cum)	Date	Title	Hiera...
1	M1	PLS	3	1	1	1	1/8/2019	Untitled	

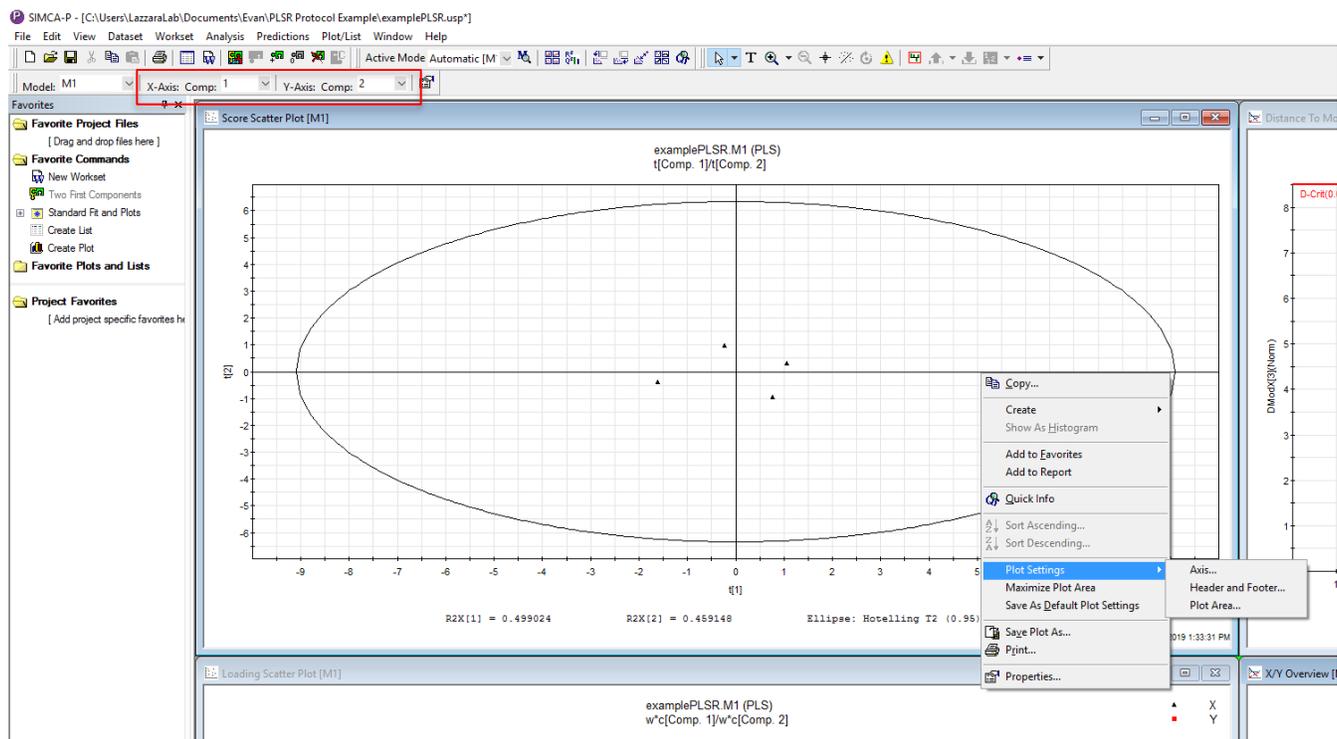
  

A	R2X	R2X(cum)	Eigenv...	R2Y	R2Y(cum)	Q2	Limit	Q2(cum)	Signifi...	It...
0	Cent.			Cent.						
1	0.499	0.499	1.5	0.694	0.694	0.0665	0.05	0.0665	R1	1
2	0.459	0.958	1.38	0.0482	0.742	-0.842	0.05	-0.0268	NS	1
3	0.0418	1	0.125	0.258	1	1	0.05	1	N3	1

- After adding sufficient components to your model, generate plots of interest by navigating to the “Analysis” tab.

The screenshot shows the SIMCA-P software interface. The 'Analysis' menu is open, showing various options. The 'Variable Importance' option is highlighted, and a sub-menu is visible with 'Plot' and 'List' options. The main window displays a loading scatter plot for 'examplePLSR.M1 (PLS)' with the x-axis labeled 't[Comp. 1]' and the y-axis labeled 't[Comp. 2]'. The plot shows a large ellipse and several data points. The status bar at the bottom indicates 'R2X[1] = 0.499024', 'R2X[2] = 0.459148', and 'Ellipse: Hotelling T2 (0.95)'. The title bar shows 'SIMCA-P 11 - 1/8/2019 1:30:28 PM'.

11. Plots can be modified by right-clicking and selecting "Plot Settings". For two-dimensional plots of parameters from models with 3 or greater components, you can select which component is on each axis by using the pull-down menus outlined below in red.



12. All plots are saved as Enhanced Metafiles (.EMF) which can be imported into another program (e.g. Adobe Illustrator) for further formatting if necessary.