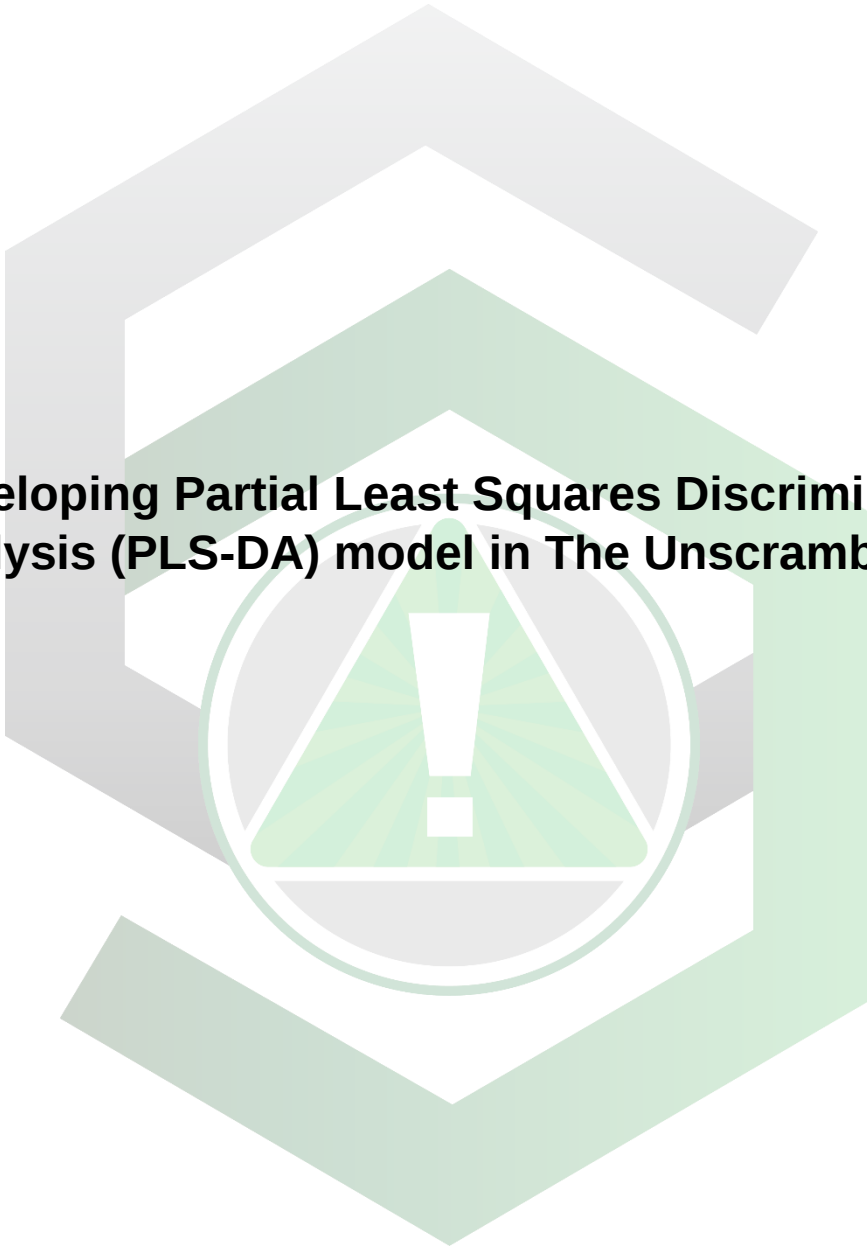


The background features a large, stylized hexagon composed of concentric, slightly offset layers in light gray and light green. In the center of this hexagon is a circular warning sign. The sign has a green border and a green background with a white exclamation mark. The text is centered over the warning sign.

## **Developing Partial Least Squares Discriminant Analysis (PLS-DA) model in The UnscramblerX**

Created by: Edwin Caballero-Agosto

Manager: Samuel Hernandez-Rivera

|                             |                                                                                  |  |                                       |
|-----------------------------|----------------------------------------------------------------------------------|--|---------------------------------------|
| SOP-01                      | Edwin Caballero-Agosto                                                           |  | University of Puerto Rico at Mayagüez |
| Effectivity:<br>Nov/12/2021 | Developing Partial Least Squares Discriminant Analysis model in The UnscramblerX |  | Revised by:                           |
| Revised:                    |                                                                                  |  | Approved by:                          |

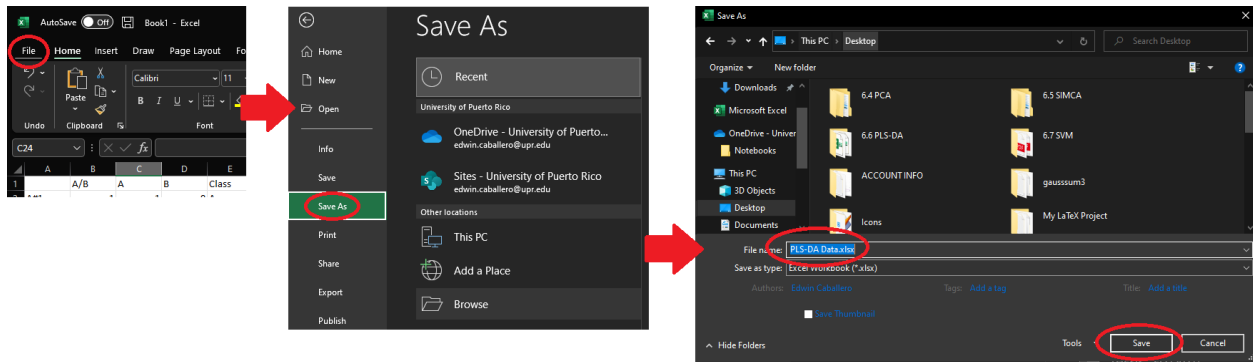
## ACQUIRING ROC CURVE DATA FROM BINARY CLASSIFIER IN MATLAB

This SOP uses the “Tutorial H” data set from The UnscramblerX program PCA tutorial.

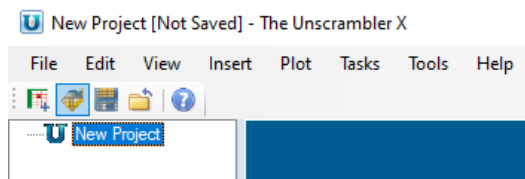
1. Create Excel file (.xlsx).
2. Write data with samples in rows and variables in columns.

|    | A    | B  | C | D     | E    | F   | G  | H   | I   | J  | K   | L  | M  | N   | O |
|----|------|----|---|-------|------|-----|----|-----|-----|----|-----|----|----|-----|---|
| 1  | A/B  | A  | B | Class | Fe   | Ti  | Ba | Ca  | K   | Mn | Rb  | Sr | Y  | Zr  |   |
| 2  | A#1  | 1  | 1 | 0 A   | 1100 | 390 | 55 | 920 | 460 | 45 | 120 | 57 | 58 | 142 |   |
| 3  | A#2  | 1  | 1 | 0 A   | 1173 | 417 | 54 | 961 | 441 | 47 | 135 | 55 | 60 | 145 |   |
| 4  | A#3  | 1  | 1 | 0 A   | 1164 | 404 | 56 | 916 | 446 | 42 | 120 | 58 | 45 | 148 |   |
| 5  | A#4  | 1  | 1 | 0 A   | 1030 | 373 | 59 | 920 | 487 | 38 | 128 | 53 | 58 | 138 |   |
| 6  | A#5  | 1  | 1 | 0 A   | 1077 | 373 | 55 | 888 | 455 | 38 | 97  | 51 | 54 | 145 |   |
| 7  | A#6  | 1  | 1 | 0 A   | 1080 | 403 | 53 | 919 | 442 | 41 | 133 | 60 | 45 | 155 |   |
| 8  | A#7  | 1  | 1 | 0 A   | 1020 | 360 | 59 | 883 | 473 | 43 | 119 | 40 | 50 | 134 |   |
| 9  | A#8  | 1  | 1 | 0 A   | 1050 | 396 | 56 | 924 | 482 | 48 | 140 | 74 | 71 | 157 |   |
| 10 | A#9  | 1  | 1 | 0 A   | 1100 | 373 | 53 | 910 | 477 | 51 | 137 | 61 | 58 | 152 |   |
| 11 | A#10 | 1  | 1 | 0 A   | 1069 | 375 | 51 | 958 | 429 | 42 | 100 | 51 | 47 | 128 |   |
| 12 | B#1  | -1 | 0 | 1 B   | 863  | 183 | 8  | 626 | 452 | 34 | 121 | 15 | 58 | 70  |   |
| 13 | B#2  | -1 | 0 | 1 B   | 1108 | 289 | 7  | 783 | 426 | 41 | 109 | 15 | 57 | 67  |   |
| 14 | B#3  | -1 | 0 | 1 B   | 1210 | 276 | 10 | 966 | 430 | 44 | 117 | 20 | 44 | 73  |   |
| 15 | B#4  | -1 | 0 | 1 B   | 1205 | 291 | 10 | 975 | 420 | 43 | 115 | 25 | 58 | 73  |   |
| 16 | B#5  | -1 | 0 | 1 B   | 1100 | 267 | 10 | 910 | 500 | 40 | 145 | 25 | 65 | 95  |   |
| 17 | B#6  | -1 | 0 | 1 B   | 1100 | 280 | 10 | 872 | 515 | 49 | 145 | 38 | 60 | 65  |   |
| 18 | B#7  | -1 | 0 | 1 B   | 689  | 114 | 9  | 534 | 404 | 26 | 110 | 25 | 50 | 55  |   |
| 19 | B#8  | -1 | 0 | 1 B   | 1186 | 257 | 10 | 940 | 431 | 40 | 121 | 20 | 53 | 73  |   |
| 20 | B#9  | -1 | 0 | 1 B   | 860  | 182 | 7  | 722 | 418 | 33 | 115 | 20 | 55 | 53  |   |
| 21 | E#1  |    |   | E     | 1050 | 195 | 46 | 865 | 400 | 36 | 97  | 10 | 50 | 160 |   |
| 22 | E#2  |    |   | E     | 1010 | 357 | 65 | 900 | 455 | 42 | 112 | 65 | 55 | 145 |   |
| 23 | E#3  |    |   | E     | 920  | 320 | 60 | 830 | 440 | 33 | 110 | 43 | 60 | 143 |   |
| 24 | E#4  |    |   | E     | 1000 | 194 | 58 | 965 | 460 | 42 | 125 | 30 | 60 | 145 |   |
| 25 | F#1  |    |   | F     | 920  | 215 | 6  | 650 | 435 | 37 | 122 | 15 | 65 | 75  |   |
| 26 | F#2  |    |   | F     | 780  | 140 | 12 | 605 | 415 | 35 | 130 | 15 | 50 | 70  |   |
| 27 | F#3  |    |   | F     | 680  | 105 | 10 | 460 | 400 | 31 | 127 | 15 | 57 | 87  |   |

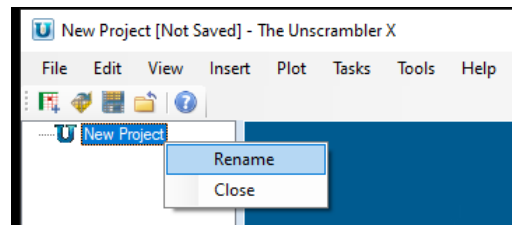
3. Save spreadsheet file as “PLS-DA Data.xlsx” in desktop by going to File → Save As... on the Desktop area.



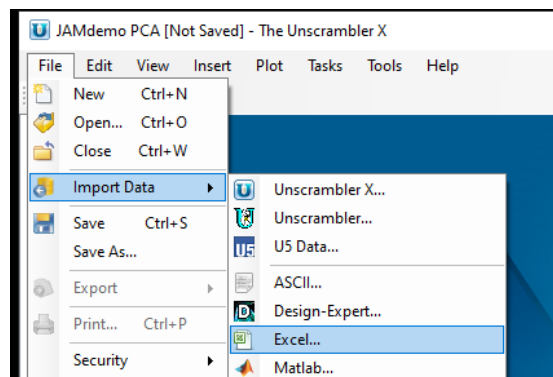
4. Open The UnscramblerX program.



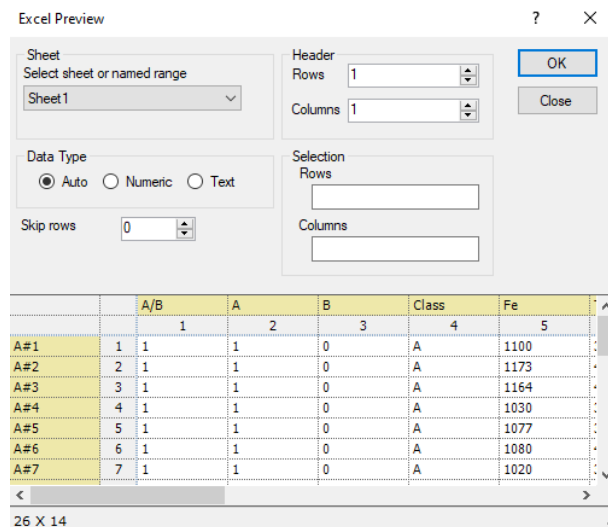
5. Rename project with right click → Rename to “PLS-DA Development”.



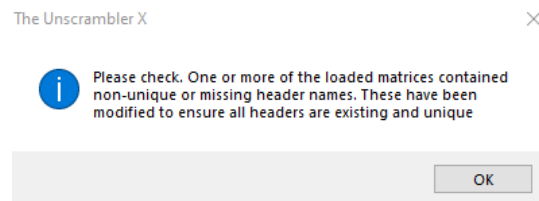
6. Import data with File → Import Data → Excel..., select the created file in the Desktop area, click “Open”.



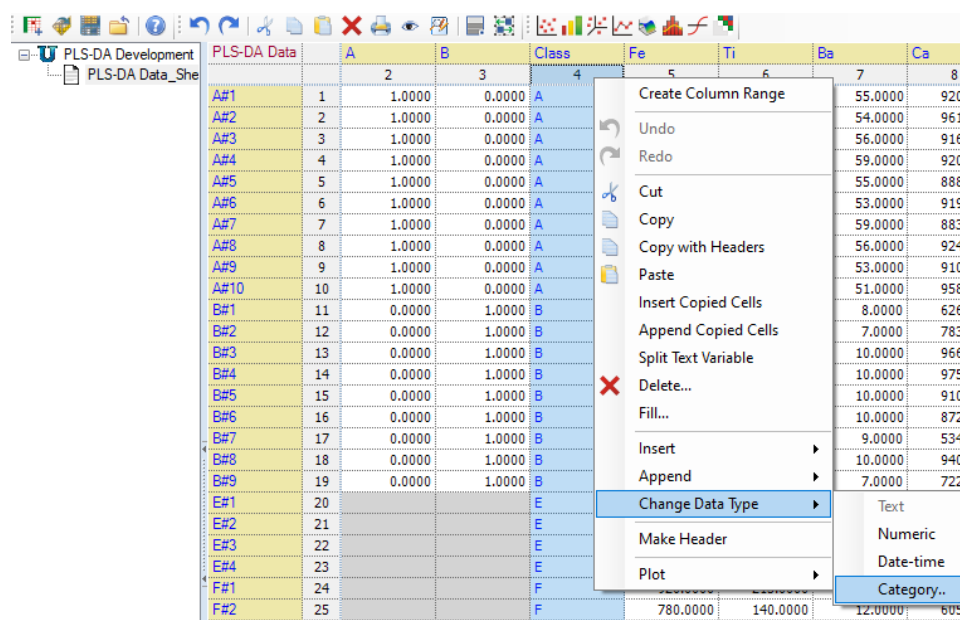
7. Select “Sheet 1” and press “OK” on the “Excel Preview” window.

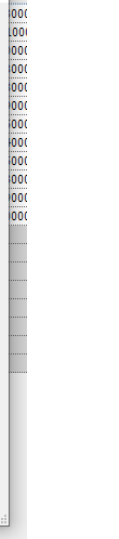


8. Click “OK” on the pop-up window if it appears.



9. Change the “Class” column from text to category by left click on the “4” below the “Class” text, right click → Change Data Type → Category...





and variable sets with Edit → Define Range...



12. Create “TRAIN” sample set by click on the first row, going to the last row with A/B values, press “Shift” tab, click on the last row, rename set to “TRAIN”, and click “Create”.

The first screenshot shows the 'Define Range' dialog box with 'Rowset' set to 1. The second screenshot shows 'Rowset' set to 1-19. The third screenshot shows the 'Rowset' renamed to 'TRAIN' and the 'Create' button circled in red.

| PLS-DA Data | A/B | A      | B      |
|-------------|-----|--------|--------|
|             | 1   | 2      |        |
| A#1         | 1   | 1.0000 | 1.0000 |
| A#2         | 2   | 1.0000 | 1.0000 |
| A#3         | 3   | 1.0000 | 1.0000 |
| A#4         | 4   | 1.0000 | 1.0000 |
| A#5         | 5   | 1.0000 | 1.0000 |
| A#6         | 6   | 1.0000 | 1.0000 |
| A#7         | 7   | 1.0000 | 1.0000 |
| A#8         | 8   | 1.0000 | 1.0000 |

| PLS-DA Data | A/B | A       | B      |
|-------------|-----|---------|--------|
|             | 1   | 2       |        |
| B#3         | 13  | -1.0000 | 0.0000 |
| B#4         | 14  | -1.0000 | 0.0000 |
| B#5         | 15  | -1.0000 | 0.0000 |
| B#6         | 16  | -1.0000 | 0.0000 |
| B#7         | 17  | -1.0000 | 0.0000 |
| B#8         | 18  | -1.0000 | 0.0000 |
| B#9         | 19  | -1.0000 | 0.0000 |
| E#1         | 20  |         |        |

13. Repeat step 12 with samples that do not have A/B values.

The first screenshot shows the 'Define Range' dialog box with 'Rowset' set to 1-19. The second screenshot shows 'Rowset' set to 20-26. The third screenshot shows the 'Rowset' renamed to 'TEST' and the 'Create' button circled in red.

| PLS-DA Data | A/B | A       | B      |
|-------------|-----|---------|--------|
|             | 1   | 2       |        |
| B#9         | 19  | -1.0000 | 0.0000 |
| E#1         | 20  |         |        |
| E#2         | 21  |         |        |
| E#3         | 22  |         |        |
| E#4         | 23  |         |        |
| F#1         | 24  |         |        |
| F#2         | 25  |         |        |
| F#3         | 26  |         |        |

| PLS-DA Data | A/B | A | B |
|-------------|-----|---|---|
|             | 1   | 2 |   |
| E#1         | 20  |   |   |
| E#2         | 21  |   |   |
| E#3         | 22  |   |   |
| E#4         | 23  |   |   |
| F#1         | 24  |   |   |
| F#2         | 25  |   |   |
| F#3         | 26  |   |   |

#### 14. Create “X” column set with column 5 to 14.

The first screenshot shows the 'Define Range' dialog with 'Row ranges' set to TRAIN (1-19) and TEST (20-26). The 'Column ranges' section is empty, and 'Columnset' is set to 5. A red arrow points to the second screenshot, which shows 'Columnset' set to 5-14. Another red arrow points to the third screenshot, which shows the 'Create' button highlighted in the 'Define Range' dialog.

| PLS-DA Data | Class | Fe | Ti        | Ba       | Ca      |
|-------------|-------|----|-----------|----------|---------|
| E#1         | 20    | E  | 1050.0000 | 195.0000 | 46.0000 |
| E#2         | 21    | E  | 1010.0000 | 357.0000 | 65.0000 |
| E#3         | 22    | E  | 920.0000  | 320.0000 | 60.0000 |
| E#4         | 23    | E  | 1000.0000 | 194.0000 | 58.0000 |
| F#1         | 24    | F  | 920.0000  | 215.0000 | 6.0000  |
| F#2         | 25    | F  | 780.0000  | 140.0000 | 12.0000 |
| F#3         | 26    | F  | 680.0000  | 105.0000 | 10.0000 |

| PLS-DA Data | Ca | K        | Mn       | Rb      | Sr       | Y       | Zr      |
|-------------|----|----------|----------|---------|----------|---------|---------|
| E#1         | 20 | 865.0000 | 400.0000 | 36.0000 | 97.0000  | 10.0000 | 50.0000 |
| E#2         | 21 | 900.0000 | 455.0000 | 42.0000 | 112.0000 | 65.0000 | 55.0000 |
| E#3         | 22 | 830.0000 | 440.0000 | 33.0000 | 110.0000 | 43.0000 | 60.0000 |
| E#4         | 23 | 965.0000 | 460.0000 | 42.0000 | 125.0000 | 30.0000 | 60.0000 |
| F#1         | 24 | 650.0000 | 435.0000 | 37.0000 | 122.0000 | 15.0000 | 65.0000 |
| F#2         | 25 | 605.0000 | 415.0000 | 35.0000 | 130.0000 | 15.0000 | 50.0000 |
| F#3         | 26 | 460.0000 | 400.0000 | 31.0000 | 127.0000 | 15.0000 | 57.0000 |

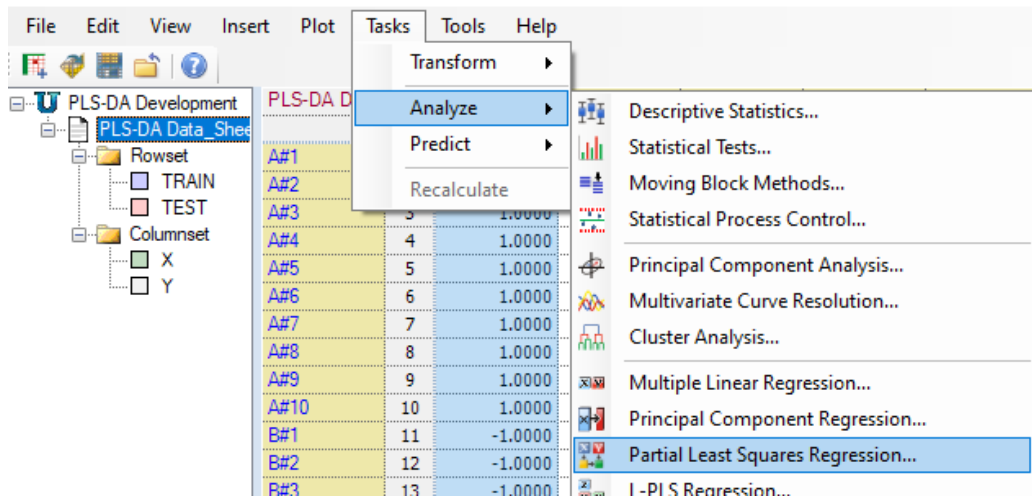
#### 15. Create “Y” column set with column 1.

The screenshot shows the 'Define Range' dialog with 'Row ranges' set to TRAIN (1-19) and TEST (20-26). The 'Column ranges' section shows 'X' (5-14) and 'Y' (1). The 'OK' button is highlighted.

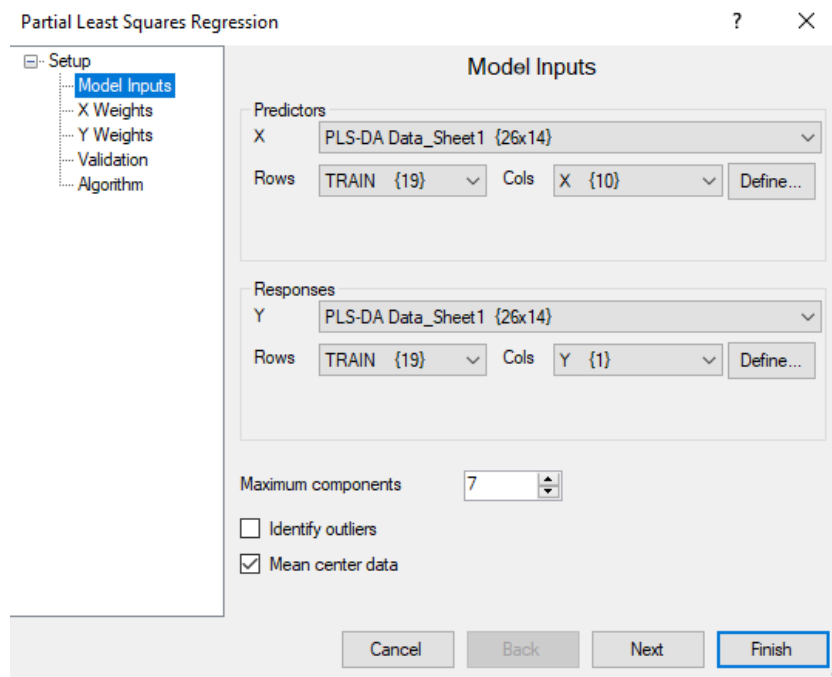
| PLS-DA Data | A/B | A | B | Class | Fe | Ti        | Ba       | Ca      |
|-------------|-----|---|---|-------|----|-----------|----------|---------|
| E#1         | 20  | 1 | 2 | 3     | 4  | 5         | 6        | 7       |
| E#2         | 21  |   |   |       |    | 1050.0000 | 195.0000 | 46.0000 |
| E#3         | 22  |   |   |       |    | 1010.0000 | 357.0000 | 65.0000 |
| E#4         | 23  |   |   |       |    | 920.0000  | 320.0000 | 60.0000 |
| F#1         | 24  |   |   |       |    | 1000.0000 | 194.0000 | 58.0000 |
| F#2         | 25  |   |   |       |    | 920.0000  | 215.0000 | 6.0000  |
| F#3         | 26  |   |   |       |    | 780.0000  | 140.0000 | 12.0000 |

#### 16. Confirm ranges by clicking “OK”. **DO NOT PRESS ENTER**, it does not save the ranges.

17. Develop PCA model with Tasks → Analyze → Partial Least Squares...



18. Select “PLS-DA Data\_Sheet 1” matrix, select “TRAIN” sample set, for the Predictors select “X” variable set, for the Responses select “Y” variable set, write “7” maximum component, and check “Mean center data” box in the “Model Inputs” tab.



19. Select “All”, select “A/(SDev+B)” in the “X Weights” tab.

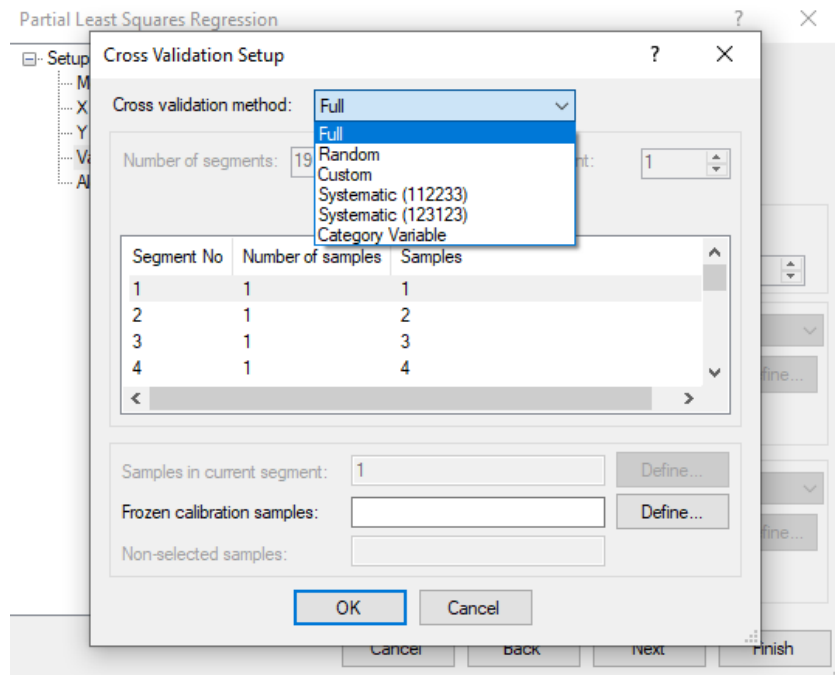
The screenshot shows the 'Partial Least Squares Regression' dialog box with the 'X Weights' tab selected. The left sidebar shows a tree view with 'Setup' expanded, containing 'Model Inputs', 'X Weights', 'Y Weights', 'Validation', and 'Algorithm'. The 'X Weights' tab contains a table with 7 rows and 2 columns: 'Variable Name' and 'Weight'. The weights are all set to  $1.00/(SDev+0.00)$ . Below the table is a dropdown menu set to '1-10' and buttons for 'Define' and 'All'. The 'Advanced' section has a 'Select' button, a 'Constant' button, a 'Downweight' button, and a 'Block weighting' button. The 'Divide blocks by SDev' checkbox is unchecked. At the bottom are 'Cancel', 'Back', 'Next', and 'Finish' buttons.

|   | Variable Name | Weight             |
|---|---------------|--------------------|
| 1 | Fe            | $1.00/(SDev+0.00)$ |
| 2 | Ti            | $1.00/(SDev+0.00)$ |
| 3 | Ba            | $1.00/(SDev+0.00)$ |
| 4 | Ca            | $1.00/(SDev+0.00)$ |
| 5 | K             | $1.00/(SDev+0.00)$ |
| 6 | Mn            | $1.00/(SDev+0.00)$ |
| 7 | Rh            | $1.00/(SDev+0.00)$ |

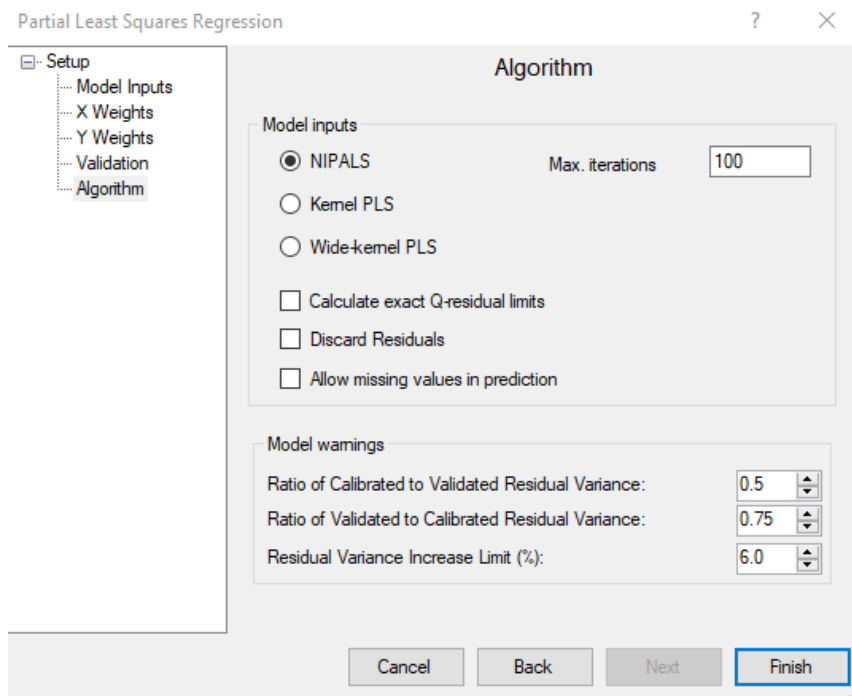
20. Select “Setup ...” button in “Validation” tab.

The screenshot shows the 'Partial Least Squares Regression' dialog box with the 'Validation' tab selected. The left sidebar shows a tree view with 'Setup' expanded, containing 'Model Inputs', 'X Weights', 'Y Weights', 'Validation', and 'Algorithm'. The 'Validation' tab has radio buttons for 'Leverage correction', 'Cross validation', and 'Test matrix'. The 'Cross validation' radio button is selected, and the 'Setup ...' button is highlighted. Below are checkboxes for 'Prediction diagnostics for CV segments' and 'Uncertainty test'. The 'Uncertainty test' section has a 'Number of components in uncertainty test' dropdown set to '1'. The 'Test matrix' section has a 'Data' dropdown set to 'X' and a 'Rows' dropdown set to 'Cols'. The 'Data' section has a 'Data' dropdown set to 'Y' and a 'Rows' dropdown set to 'Cols'. At the bottom are 'Cancel', 'Back', 'Next', and 'Finish' buttons.

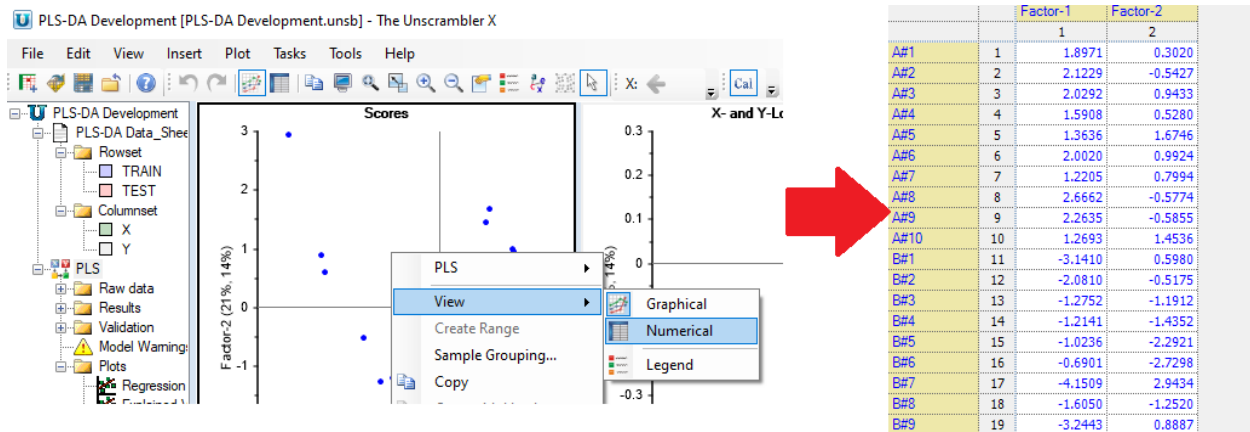
21. Select “Full” cross validation method in the “Cross Validation Setup” window, and press “OK”.



22. Select “NIPALS” circle in the “Algorithm” tab and click “Finish”.



23. Copy data for a plot by right clicking on the graph → View → Numerical.



24. Save file with File → Save As...

## Appendix Section

Category columns describing data, Response matrix, and Predictor matrix with both TRAIN and TEST sample sets.

|      | A/B | A | B | Class | Fe   | Ti  | Ba | Ca  | K   | Mn | Rb  | Sr | Y  | Zr  |
|------|-----|---|---|-------|------|-----|----|-----|-----|----|-----|----|----|-----|
| A#1  | 1   | 1 | 0 | A     | 1100 | 390 | 55 | 920 | 460 | 45 | 120 | 57 | 58 | 142 |
| A#2  | 1   | 1 | 0 | A     | 1173 | 417 | 54 | 961 | 441 | 47 | 135 | 55 | 60 | 145 |
| A#3  | 1   | 1 | 0 | A     | 1164 | 404 | 56 | 916 | 446 | 42 | 120 | 58 | 45 | 148 |
| A#4  | 1   | 1 | 0 | A     | 1030 | 373 | 59 | 920 | 487 | 38 | 128 | 53 | 58 | 138 |
| A#5  | 1   | 1 | 0 | A     | 1077 | 373 | 55 | 888 | 455 | 38 | 97  | 51 | 54 | 145 |
| A#6  | 1   | 1 | 0 | A     | 1080 | 403 | 53 | 919 | 442 | 41 | 133 | 60 | 45 | 155 |
| A#7  | 1   | 1 | 0 | A     | 1020 | 360 | 59 | 883 | 473 | 43 | 119 | 40 | 50 | 134 |
| A#8  | 1   | 1 | 0 | A     | 1050 | 396 | 56 | 924 | 482 | 48 | 140 | 74 | 71 | 157 |
| A#9  | 1   | 1 | 0 | A     | 1100 | 373 | 53 | 910 | 477 | 51 | 137 | 61 | 58 | 152 |
| A#10 | 1   | 1 | 0 | A     | 1069 | 375 | 51 | 958 | 429 | 42 | 100 | 51 | 47 | 128 |
| B#1  | -1  | 0 | 1 | B     | 863  | 183 | 8  | 626 | 452 | 34 | 121 | 15 | 58 | 70  |
| B#2  | -1  | 0 | 1 | B     | 1108 | 289 | 7  | 783 | 426 | 41 | 109 | 15 | 57 | 67  |
| B#3  | -1  | 0 | 1 | B     | 1210 | 276 | 10 | 966 | 430 | 44 | 117 | 20 | 44 | 73  |
| B#4  | -1  | 0 | 1 | B     | 1205 | 291 | 10 | 975 | 420 | 43 | 115 | 25 | 58 | 73  |
| B#5  | -1  | 0 | 1 | B     | 1100 | 267 | 10 | 910 | 500 | 40 | 145 | 25 | 65 | 95  |
| B#6  | -1  | 0 | 1 | B     | 1100 | 280 | 10 | 872 | 515 | 49 | 145 | 38 | 60 | 65  |
| B#7  | -1  | 0 | 1 | B     | 689  | 114 | 9  | 534 | 404 | 26 | 110 | 25 | 50 | 55  |
| B#8  | -1  | 0 | 1 | B     | 1186 | 257 | 10 | 940 | 431 | 40 | 121 | 20 | 53 | 73  |
| B#9  | -1  | 0 | 1 | B     | 860  | 182 | 7  | 722 | 418 | 33 | 115 | 20 | 55 | 53  |
| E#1  |     |   |   | E     | 1050 | 195 | 46 | 865 | 400 | 36 | 97  | 10 | 50 | 160 |
| E#2  |     |   |   | E     | 1010 | 357 | 65 | 900 | 455 | 42 | 112 | 65 | 55 | 145 |
| E#3  |     |   |   | E     | 920  | 320 | 60 | 830 | 440 | 33 | 110 | 43 | 60 | 143 |
| E#4  |     |   |   | E     | 1000 | 194 | 58 | 965 | 460 | 42 | 125 | 30 | 60 | 145 |
| F#1  |     |   |   | F     | 920  | 215 | 6  | 650 | 435 | 37 | 122 | 15 | 65 | 75  |
| F#2  |     |   |   | F     | 780  | 140 | 12 | 605 | 415 | 35 | 130 | 15 | 50 | 70  |
| F#3  |     |   |   | F     | 680  | 105 | 10 | 460 | 400 | 31 | 127 | 15 | 57 | 87  |

Advisor Signature

---

Co-Advisor Signature

---