Developing Principal Component Analysis (PCA) 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: Nov/11/2021	Developing Principal Component Analysis model in The UnscramblerX		Revised by:
Revised:			Approved by:

ACQUIRING ROC CURVE DATA FROM BINARY CLASSIFIER IN MATLAB

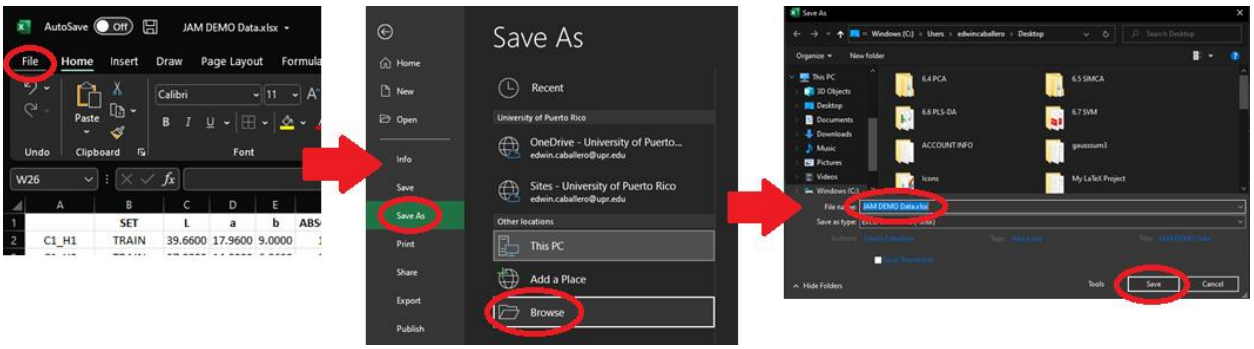
This SOP uses the "JAMdemo" data from The UnscramblerX program PCA tutorial.

1. Create Excel file (.xlsx).

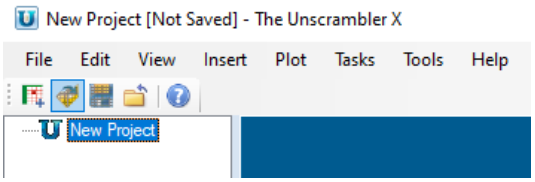
2. Write data with samples in rows and variables in columns.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
1		SET	CULTIVAR	HARVEST TIME	L	a	b	PREFERENCE	ABSORBANCE	SOLUBLE	ACIDITY	REDNESS	COLOUR	SHININESS	R.SMELL	R.FLAV	SWEETNESS	SOURNESS	BITTERNESS	OFF-FLAV	JUICINESS	THICKNESS	CHEW.RES
2	C1_H1	TRAIN	C1	H1	39.6600	17.9600	9.0000	4.3000	1.8360	8.3000	16.0200	3.9167	4.4444	3.8333	3.7500	3.8889	5.1389	3.8333	3.0556	3.9167	5.4444	4.1667	3.8056
3	C1_H2	TRAIN	C1	H2	37.8200	14.8000	6.8600	7.6000	2.2050	10.1000	17.0400	6.1667	6.0556	4.2778	3.7222	3.6944	5.5833	3.5000	2.9167	4.3056	6.0278	4.1389	3.4167
4	C1_H3	TRAIN	C1	H3	34.5600	7.8000	3.5800	8.5000	5.2370	8.0000	25.9300	8.4444	8.0278	5.1111	3.8333	3.6389	4.7778	4.4167	3.6667	3.9444	6.0278	3.1944	4.2778
5	C2_H1	TRAIN	C2	H1	39.1400	18.0400	9.1400	4.0000	1.8150	7.8000	25.5300	3.8889	4.3333	4.2778	4.3889	4.3611	4.6944	4.1944	2.9167	3.1944	5.3333	4.1111	3.8333
6	C2_H2	TRAIN	C2	H2	36.8200	14.3000	6.7400	3.0000	2.6730	8.8000	32.7100	4.9722	5.6111	4.7778	4.7778	3.9444	3.6944	4.4444	3.3889	3.8889	6.0833	3.7500	3.7778
7	C2_H3	TRAIN	C2	H3	34.8800	8.3600	3.9000	8.0000	4.6050	6.9000	21.5800	7.7500	7.6389	5.0278	4.0556	3.8056	4.3889	4.3333	3.4722	3.8333	6.0278	3.3889	4.4444
8	C3_H1	TRAIN	C3	H1	39.0400	18.9200	9.4000	1.8000	2.1720	8.5000	39.1200	3.8889	4.8611	4.7500	4.6667	4.0000	3.1389	4.4722	3.6944	3.5278	5.3056	4.8056	4.5278
9	C3_H2	TRAIN	C3	H2	36.5200	13.3600	6.6000	5.9000	2.7490	8.4000	22.0300	6.0278	6.4444	5.1667	5.5556	5.3889	4.6944	4.7222	2.4722	2.7500	6.2222	3.6667	3.1944
10	C3_H3	TRAIN	C3	H3	37.1200	15.0000	7.0400	5.0000	3.4130	7.5000	33.7700	6.6389	7.0833	4.9722	5.4444	4.9167	3.8056	4.9167	3.5000	2.6944	6.0556	3.5278	3.8889
11	C4_H1	TRAIN	C4	H1	38.4000	17.2400	8.2600	3.2000	2.1070	9.3000	29.5000	4.4444	4.9722	4.6667	4.9167	4.6389	4.3056	4.3056	3.2222	2.9722	5.7222	4.3056	3.8333
12	C4_H2	TRAIN	C4	H2	37.7600	16.4000	7.9000	2.9000	2.7320	7.9000	32.8400	5.1389	5.8333	5.4444	5.6111	5.3056	3.5833	4.7222	3.2500	2.6389	6.1389	4.2222	3.7500
13	C4_H3	TRAIN	C4	H3	36.7800	13.8000	6.3000	6.0000	2.7580	9.0000	21.4900	5.9167	6.3889	5.0833	5.8611	5.7500	4.9722	5.2222	2.3611	2.1389	6.5278	3.6944	3.3333
14	Unknown#1	TEST										5.9238	6.5768	4.9118	5.6358	5.3507	4.9558	4.7141	2.1761	2.2950	6.4443	3.6877	3.2514
15	Unknown#2	TEST										3.6900	4.5249	3.5321	3.9864	4.3872	4.8710	4.2221	3.4792	4.1784	5.7977	3.6106	3.5903
16	Unknown#3	TEST										5.0692	5.6856	5.3573	5.6122	5.7164	3.5544	5.0140	2.7983	2.2705	6.1268	4.1956	3.6989
17	Unknown#4	TEST										8.1095	7.4557	5.1548	4.0683	3.8000	4.3821	4.3804	3.9409	3.7002	5.3857	3.0816	4.3668
18	Unknown#5	TEST										4.4446	4.9435	5.4550	4.9007	4.5559	4.1389	4.9485	3.6582	2.8365	5.6344	4.6876	3.1538
19	Unknown#6	TEST										3.9933	4.3199	4.0366	4.4630	4.5905	4.7935	3.9779	2.5134	3.5143	5.3775	3.9395	4.2298
20	Unknown#7	TEST										3.7914	4.8963	3.8760	4.8020	3.6254	3.1181	4.3041	3.8387	3.5680	5.0389	5.3752	5.1370
21	Unknown#8	TEST										6.5477	7.2265	5.0024	5.5660	4.6311	4.1463	4.6890	3.9648	2.6621	6.4781	3.1055	4.3307

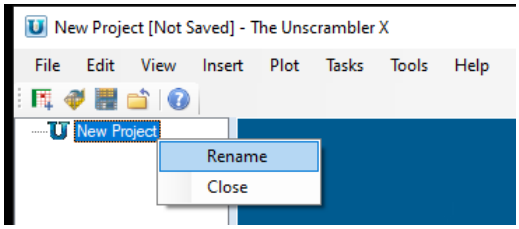
3. Save spreadsheet file as "JAMdemo.xlsxt" in desktop by going to File → Save As… on the Desktop area.



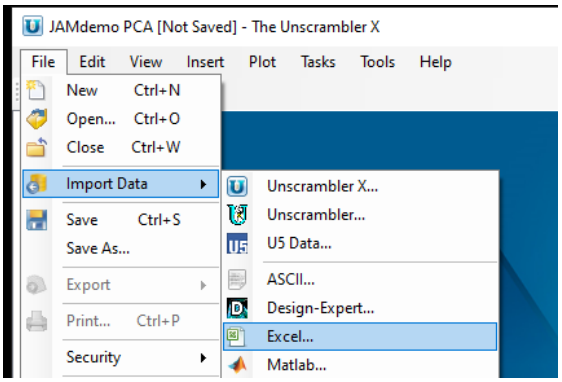
4. Open The UnscramblerX program.



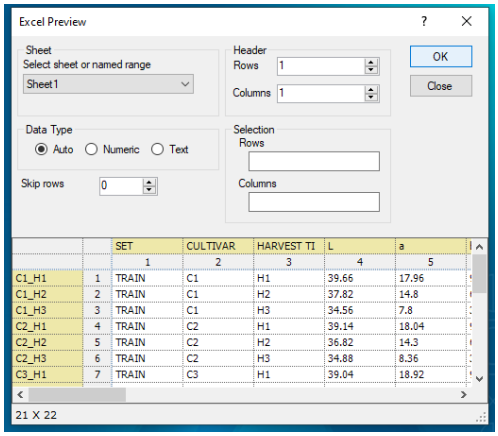
5. Rename project with right click → Rename to “JAMdemo PCA”.



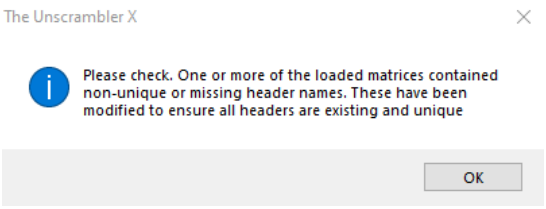
6. Import data with File → Import Data → Excel…, select the created file in the Desktop area, click “Open”.



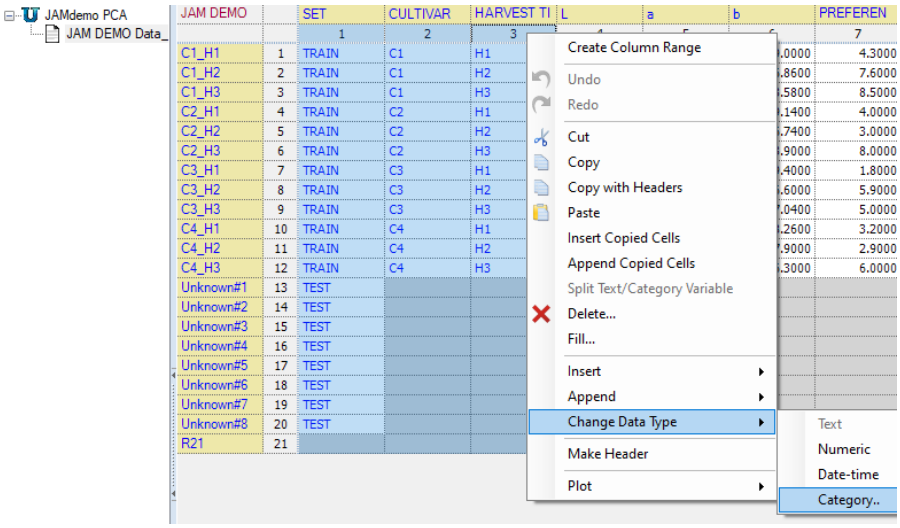
7. Select “Sheet 1” and press “OK” on the “Excel Preview” window.



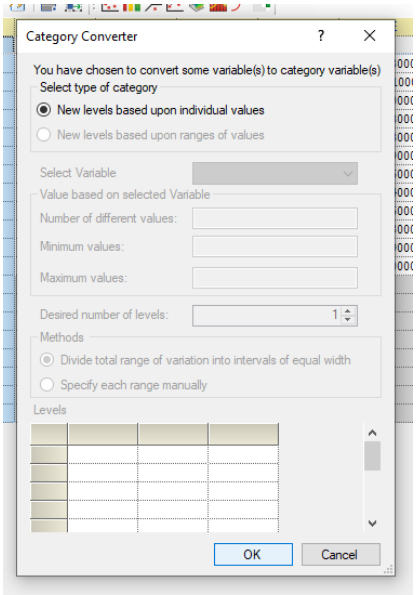
8. Click “OK” on the pop-up window.



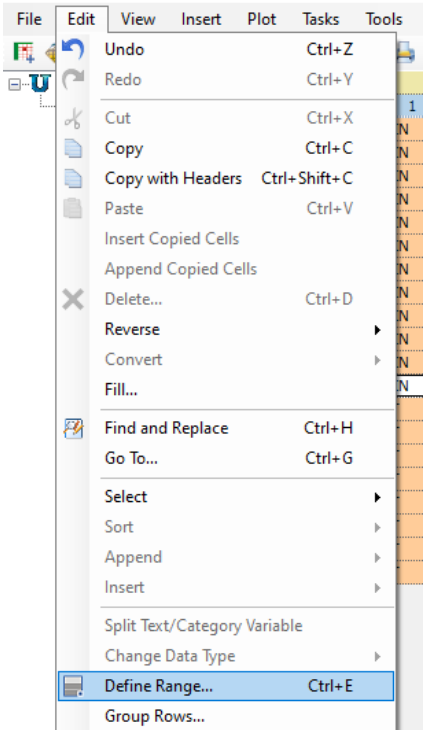
9. Change the “SET”, “CULTIVAR”, and “HARVEST TIME” columns from text to category by left click on the “1” below the “SET” text, press “Shift” tab, left click on the “3” below “HARVEST TIME”, right click → Change Data Type → Category…



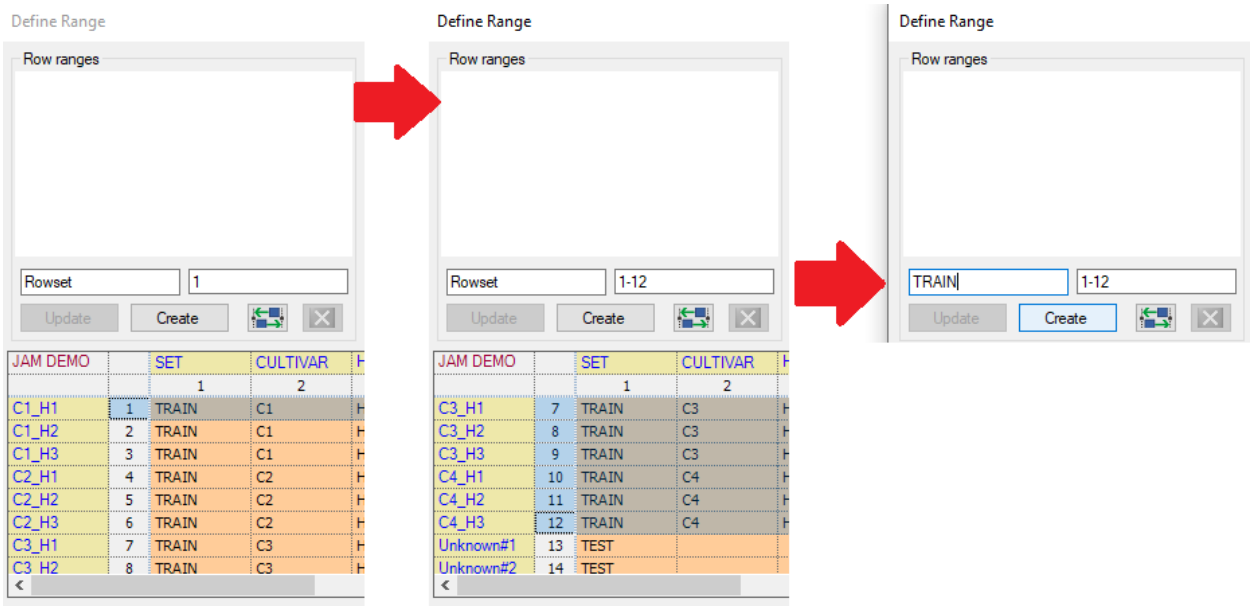
10. Click “OK” on the “Category Converter” window.



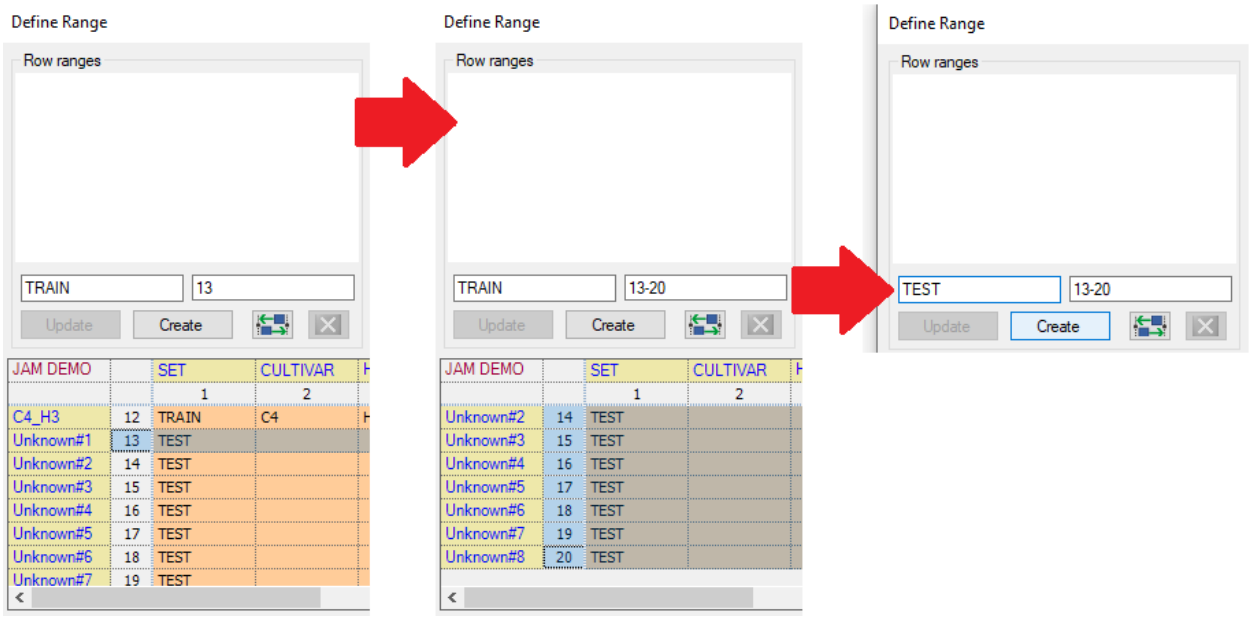
11. Create sample and variable sets with Edit → Define Range…



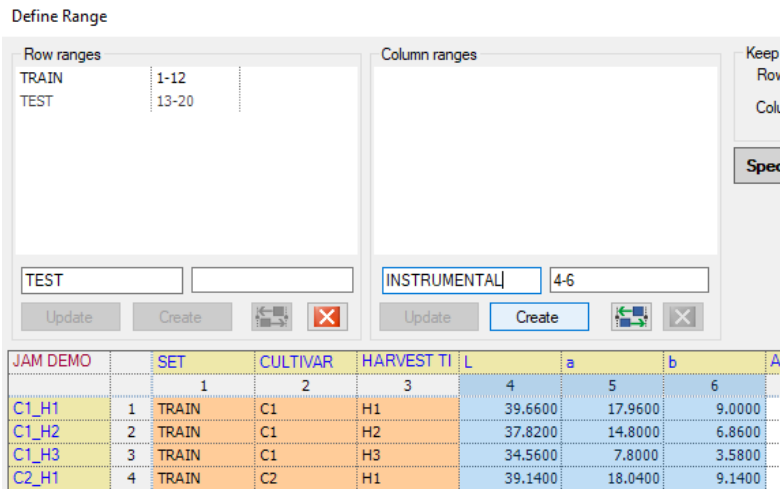
12. Create “TRAIN” sample set by click on the first row, going to the last row of the “TRAIN” samples, press “Shift” tab, click on the last row, rename set to “TRAIN”, and click “Create”.



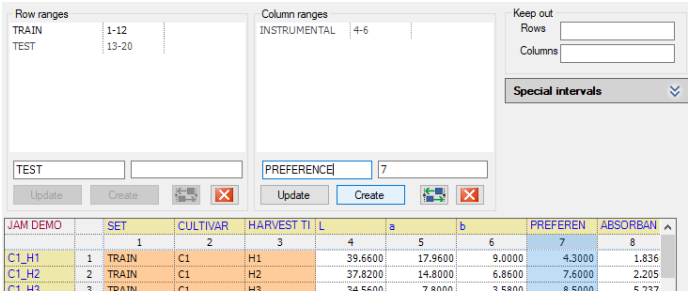
13. Repeat step 12 with “TEST” sample set.



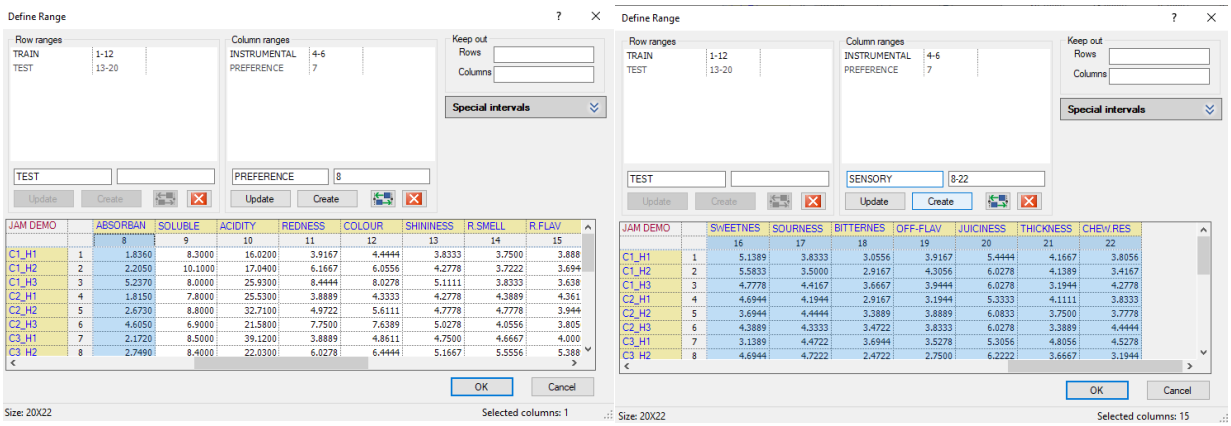
14. Create “Instrumental” column set with column 4, 5, and 6.



15. Create “PREFERENCE” variable set with column 7.

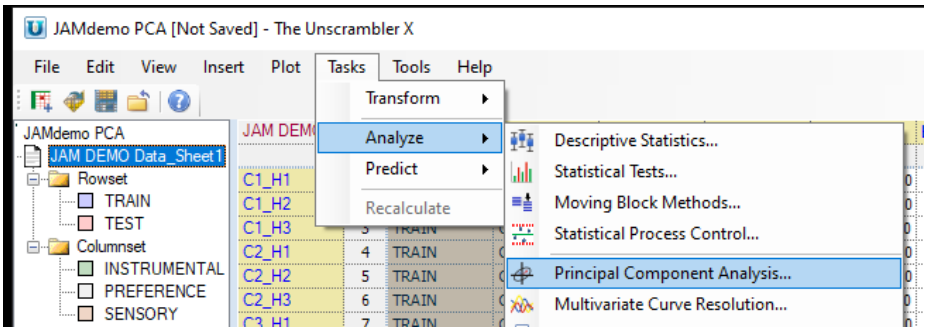


16. Create “SENSORY” variable set with column 8 to 22.

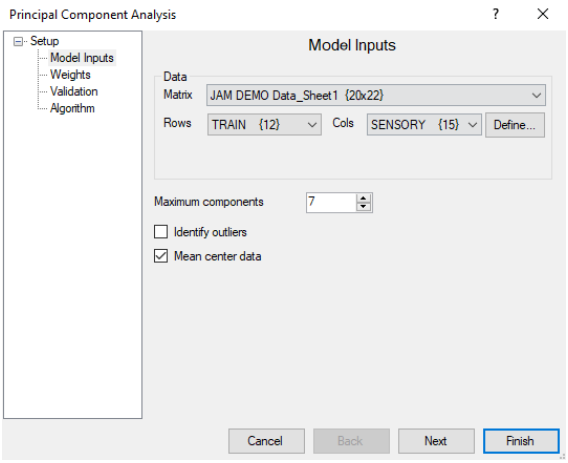


17. Confirm ranges by clicking “OK”. DO NOT PRESS ENTER, it does not save the ranges.

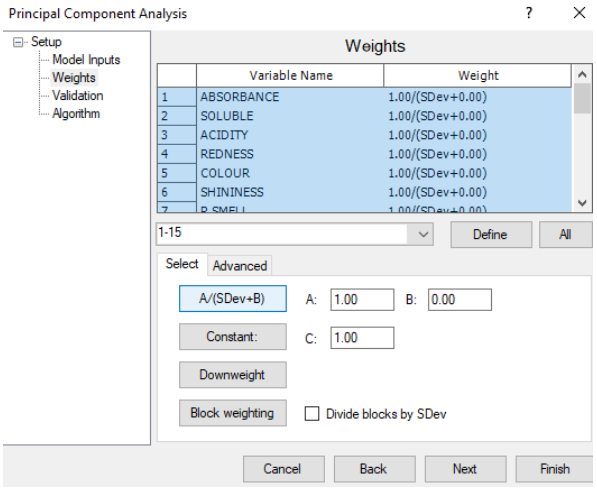
18. Develop PCA model with Tasks → Analyze → Principal Component Analysis…



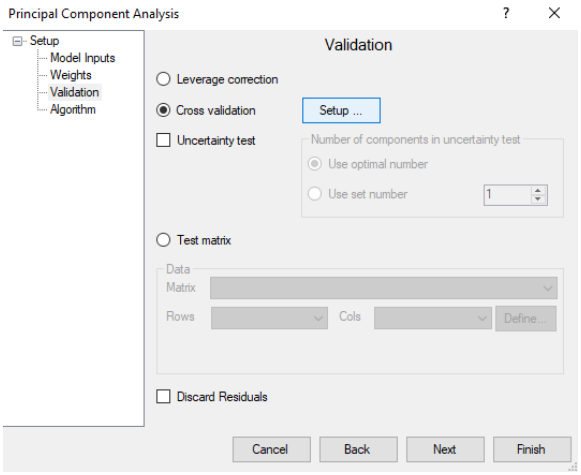
19. Select “JAM DEMO Data_Sheet1” matrix, select “TRAIN” sample set, select “SENSORY” variable set, write “7” maximum component, and check “Mean center data” box in the “Model Inputs” tab.



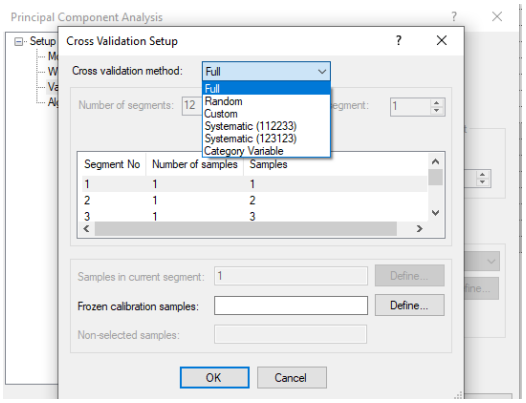
20. Select “All”, select “A/(SDev+B)” in the “Weights” tab.



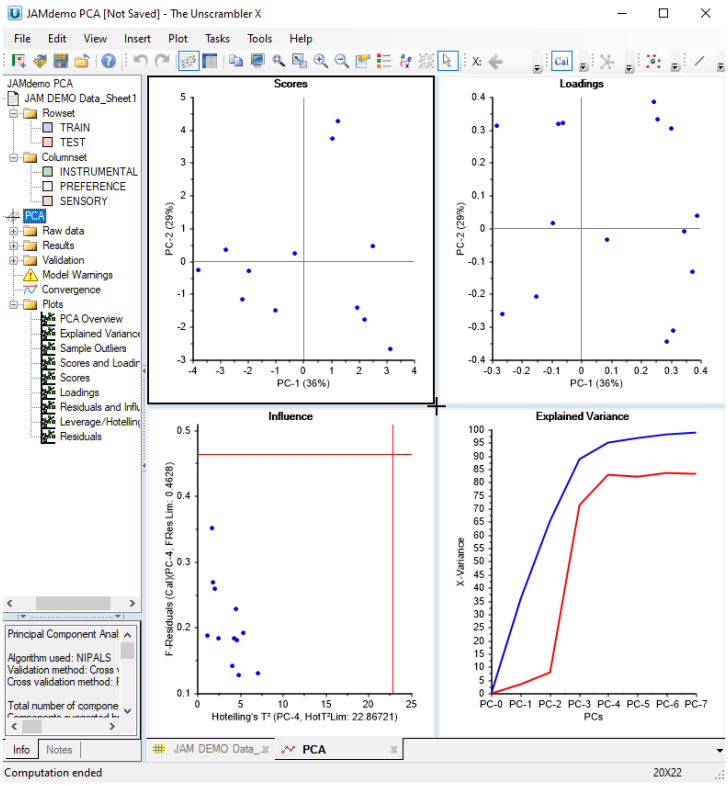
21. Select “Setup …” button in “Validation” tab.



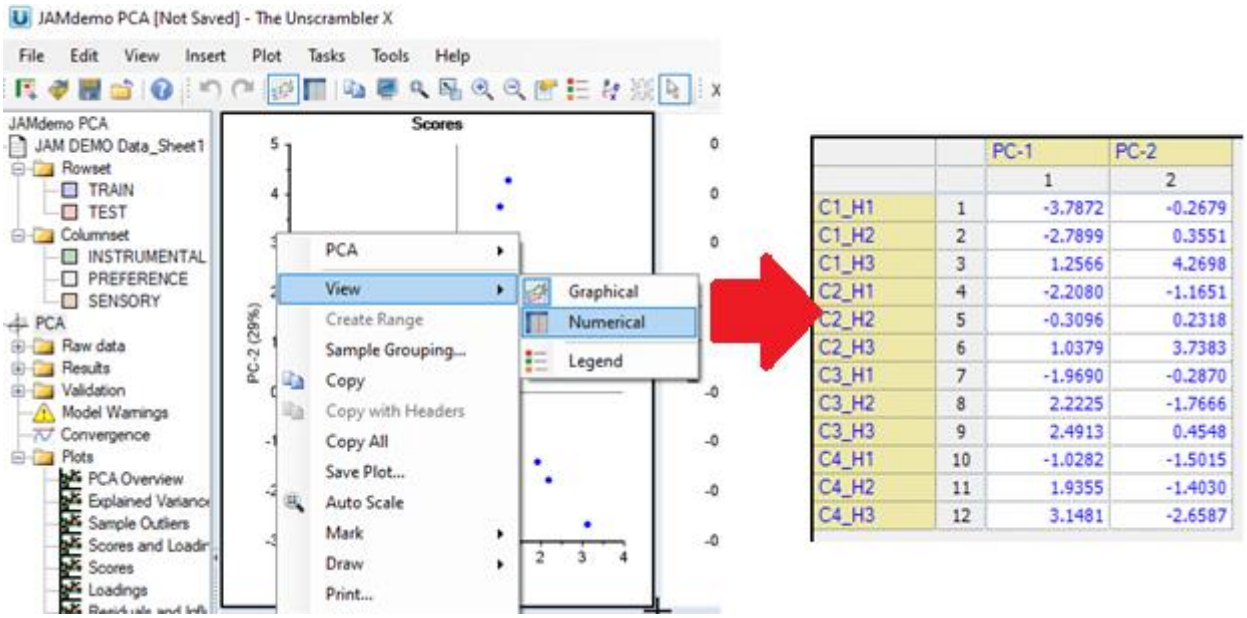
22. Select “Full” cross validation method in the “Cross Validation Setup” window, and press “OK”.



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



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



		PC-1	PC-2
		1	2
C1_H1	1	-3.7872	-0.2679
C1_H2	2	-2.7899	0.3551
C1_H3	3	1.2566	4.2698
C2_H1	4	-2.2080	-1.1651
C2_H2	5	-0.3096	0.2318
C2_H3	6	1.0379	3.7383
C3_H1	7	-1.9690	-0.2870
C3_H2	8	2.2225	-1.7666
C3_H3	9	2.4913	0.4548
C4_H1	10	-1.0282	-1.5015
C4_H2	11	1.9355	-1.4030
C4_H3	12	3.1481	-2.6587

25. Save file with File → Save As…

Appendix Section

Category columns describing data, instrumental data, and Preference predictor variable with both TRAIN and TEST sample sets.

	SET	CULTIVAR	HARVEST TIME	L	a	b	PREFERENCE
C1_H1	TRAIN	C1	H1	39.6600	17.9600	9.0000	4.3000
C1_H2	TRAIN	C1	H2	37.8200	14.8000	6.8600	7.6000
C1_H3	TRAIN	C1	H3	34.5600	7.8000	3.5800	8.5000
C2_H1	TRAIN	C2	H1	39.1400	18.0400	9.1400	4.0000
C2_H2	TRAIN	C2	H2	36.8200	14.3000	6.7400	3.0000
C2_H3	TRAIN	C2	H3	34.8800	8.3600	3.9000	8.0000
C3_H1	TRAIN	C3	H1	39.0400	18.9200	9.4000	1.8000
C3_H2	TRAIN	C3	H2	36.5200	13.3600	6.6000	5.9000
C3_H3	TRAIN	C3	H3	37.1200	15.0000	7.0400	5.0000
C4_H1	TRAIN	C4	H1	38.4000	17.2400	8.2600	3.2000
C4_H2	TRAIN	C4	H2	37.7600	16.4000	7.9000	2.9000
C4_H3	TRAIN	C4	H3	36.7800	13.8000	6.3000	6.0000
Unknown#1	TEST						
Unknown#2	TEST						
Unknown#3	TEST						
Unknown#4	TEST						
Unknown#5	TEST						
Unknown#6	TEST						
Unknown#7	TEST						
Unknown#8	TEST						

SENSORY variable set with for TRAIN and TEST sample sets.

ABSORBANCE	SOLUBLE	ACIDITY	REDNESS	COLOUR	SHININESS	R.SMELL	R.FLAV	SWEETNESS	SOURNESS	BITTERNESS	OFF-FLAV	JUICINESS	THICKNESS	CHEW.RES
1.8360	8.3000	16.0200	3.9167	4.4444	3.8333	3.7500	3.8889	5.1389	3.8333	3.0556	3.9167	5.4444	4.1667	3.8056
2.2050	10.1000	17.0400	6.1667	6.0556	4.2778	3.7222	3.6944	5.5833	3.5000	2.9167	4.3056	6.0278	4.1389	3.4167
5.2370	8.0000	25.9300	8.4444	8.0278	5.1111	3.8333	3.6389	4.7778	4.4167	3.6667	3.9444	6.0278	3.1944	4.2778
1.8150	7.8000	25.5300	3.8889	4.3333	4.2778	4.3889	4.3611	4.6944	4.1944	2.9167	3.1944	5.3333	4.1111	3.8333
2.6730	8.8000	32.7100	4.9722	5.6111	4.7778	4.7778	3.9444	3.6944	4.4444	3.3889	3.8889	6.0833	3.7500	3.7778
4.6050	6.9000	21.5800	7.7500	7.6389	5.0278	4.0556	3.8056	4.3889	4.3333	3.4722	3.8333	6.0278	3.3889	4.4444
2.1720	8.5000	39.1200	3.8889	4.8611	4.7500	4.6667	4.0000	3.1389	4.4722	3.6944	3.5278	5.3056	4.8056	4.5278
2.7490	8.4000	22.0300	6.0278	6.4444	5.1667	5.5556	5.3889	4.6944	4.7222	2.4722	2.7500	6.2222	3.6667	3.1944
3.4130	7.5000	33.7700	6.6389	7.0833	4.9722	5.4444	4.9167	3.8056	4.9167	3.5000	2.6944	6.0556	3.5278	3.8889
2.1070	9.3000	29.5000	4.4444	4.9722	4.6667	4.9167	4.6389	4.3056	4.3056	3.2222	2.9722	5.7222	4.3056	3.8333
2.7320	7.9000	32.8400	5.1389	5.8333	5.4444	5.6111	5.3056	3.5833	4.7222	3.2500	2.6389	6.1389	4.2222	3.7500
2.7580	9.0000	21.4900	5.9167	6.3889	5.0833	5.8611	5.7500	4.9722	5.2222	2.3611	2.1389	6.5278	3.6944	3.3333
			5.9238	6.5768	4.9118	5.6358	5.3507	4.9558	4.7141	2.1761	2.2950	6.4443	3.6877	3.2514
			3.6900	4.5249	3.5321	3.9864	4.3872	4.8710	4.2221	3.4792	4.1784	5.7977	3.6106	3.5903
			5.0692	5.6856	5.3573	5.6122	5.7164	3.3544	5.0140	2.7983	2.2705	6.1268	4.1956	3.6989
			8.1095	7.4557	5.1548	4.0683	3.8000	4.3821	4.3804	3.9409	3.7002	5.3857	3.0816	4.3668
			4.4446	4.9435	5.4550	4.9007	4.5559	4.1389	4.9485	3.6582	2.8365	5.6344	4.6876	3.1538
			3.9933	4.3199	4.0366	4.4630	4.5905	4.7935	3.9779	2.5134	3.5143	5.3775	3.9395	4.2298
			3.7914	4.8963	3.8760	4.8020	3.6254	3.1181	4.3041	3.8387	3.5680	5.0389	5.3752	5.1370
			6.5477	7.2265	5.0024	5.5660	4.6311	4.1463	4.6890	3.9648	2.6621	6.4781	3.1055	4.3307

Advisor Signature

Co-Advisor Signature