

การใช้โปรแกรม The Unscrambler X 10.5 ในการทำ

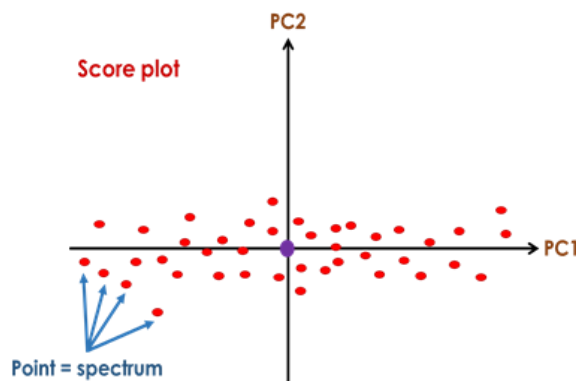
Principal component analysis (PCA)

The Unscrambler X 10.5 เป็นโปรแกรมที่ใช้ในการวิเคราะห์ข้อมูลหลายตัวแปร (Multivariate data analysis) เพื่อเข้าใจความสัมพันธ์ระหว่างตัวแปรในชุดข้อมูลขนาดใหญ่ ซึ่งเป็นวิธีที่ง่าย แม่นยำ และรวดเร็ว โดยข้อมูลที่สามารถนำเอาเข้ามาวิเคราะห์ด้วยโปรแกรมนี้ สามารถเป็นได้ทั้งข้อมูล spectrum หรือข้อมูลตัวเลขที่ได้จาก chemical และ physical โดยรายงานฉบับนี้จะกล่าวถึง ขั้นตอนในการทำ data-preprocessing และการทำ PCA-score plot

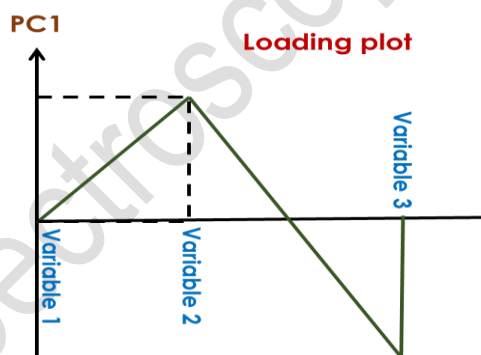
Principal component analysis (PCA) เป็นการวิเคราะห์ข้อมูลที่เหมาะสมกับกลุ่มตัวอย่าง (data set) ขนาดใหญ่ และในข้อมูลตัวอย่างนั้นมีรายละเอียดของข้อมูลอยู่จำนวนมาก การทำ PCA จะช่วยให้สามารถแยกความแตกต่างจากตัวอย่างหนึ่งออกจากอีกตัวอย่างหนึ่งได้โดยอาศัยตัวแปรที่สัมพันธ์กันในกลุ่มตัวอย่าง ซึ่งจะช่วยให้เป็นพื้นฐานในการวิเคราะห์ขั้นสูงหรือการวิเคราะห์เพื่อจำแนกประเภท เช่นการทำ SIMCA, PLS-R หรือ PLS-DA ได้ โดยในคู่มือการใช้งานฉบับนี้ จะกล่าวถึงการวิเคราะห์ตัวอย่างจากข้อมูล spectrum โดยการทำ PCA เพื่อแยกความแตกต่างของกลุ่มตัวอย่าง โดยข้อมูลที่นำมาวิเคราะห์ประกอบด้วย

	Variable 1	Variable 2	Variable 3
Object 1			
Object 2	X1	X2	X3
Object 3			
Object 4			

- Spectrum ซึ่งแทนด้วย object
- Variable ซึ่งแทนด้วย frequency (wavenumber) ซึ่ง frequency ที่เราสนใจใช้วิเคราะห์ในครั้งนี้ คือช่วง $4000-400\text{ cm}^{-1}$



Score plot จะใช้ในการอธิบายว่าตัวอย่างที่นำเข้ามาวิเคราะห์ต่างหรือไม่ต่างกัน จะประกอบด้วย point ซึ่งหมายถึง spectrum และกลุ่มของ spectrum ที่กระจายตัวอยู่เรียกว่า swarm of spectrum โดยแกน PC ที่ลากผ่านกลุ่ม swarm of spectrum มากที่สุด จะได้เป็น PC1 และ PCA score plot ที่ได้ จะสัมพันธ์กับ variable หรือ frequency ที่ให้ผลดีที่สุดต่อการแยก เมื่อได้แกน PC2 แล้วจากนั้นบริเวณจุด center of mean scattering ในแนวตั้งฉากกับ PC1 จะได้เป็น PC2 ซึ่งจะช่วยในการอธิบายว่า variable ที่มีอิทธิพลถัดมาเพื่อช่วยในการแยก ในกรณีที่ spectrum มี error หรือมี scattering สูง อาจต้องใช้ PC3, PC4... เพื่อช่วยในการแยก ตามลำดับ

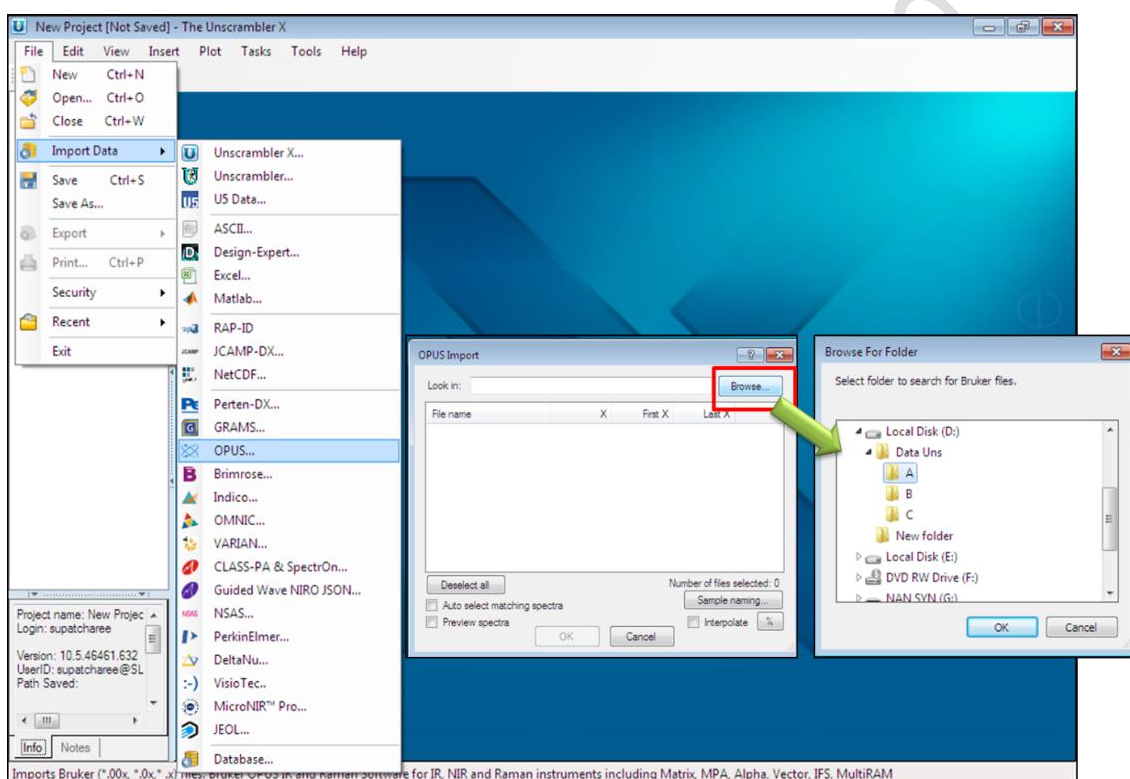


Loading plot ใช้ในการอธิบายกลุ่มของ spectrum ใน PCA score plot โดยจะเป็นตัวบอกว่า variable หรือ frequency ใดที่มีความสัมพันธ์กันและให้ผลดีที่สุดต่อการแยก โดยค่า loading ที่สูงจะสัมพันธ์กับ variable (frequency) ใน PCA score plot ซึ่งการอ่านผล loading จะสัมพันธ์กับ PCA score plot และจะต้องใช้ในการอ่านผลคู่กันเสมอ โดยพิจารณาจากแกนค่าบวกและลบในการแยก ข้อมูลที่ผ่านการทำ preprocessing แบบไม่ derivative จะอ่านผล loading แกนค่าบวกสัมพันธ์กับแกน PCA score plot ค่าบวก และ loading แกนค่าลบสัมพันธ์กับแกน PCA score plot ค่าลบ

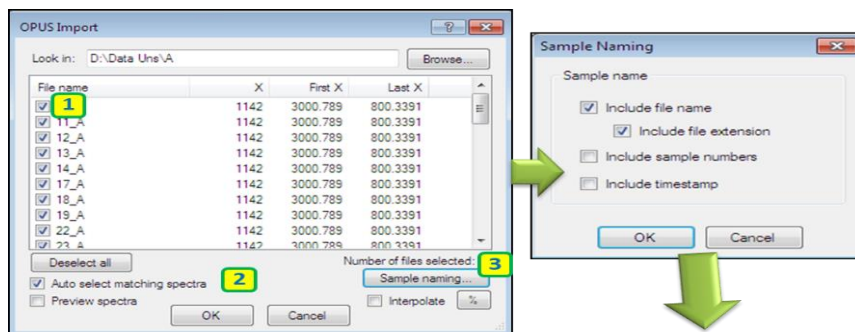
ในเอกสารคู่มือฉบับนี้จะอธิบายการทำ PCA ของตัวอย่าง 2 กลุ่ม ได้แก่ A และ C

การทำ PCA จาก 2 ตัวอย่าง

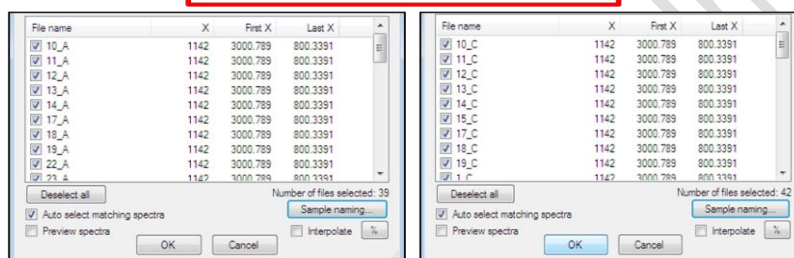
1. Import ข้อมูล spectrum ของแต่ละกลุ่มตัวอย่างเข้ามา (ในตัวอย่างประกอบด้วยข้อมูลกลุ่ม A และ C)
 - a. เลือกไฟล์ Import data
 - b. เลือก OPUS จะปรากฏหน้าต่าง OPUS Import
 - c. เลือก Browse เพื่อ import สเปคตรัมเข้ามา
 - d. เลือก OK



- e. เลือกข้อมูลสเปคตรัมโดยคลิกที่ File name (1)
- f. เลือก Auto select matching spectra (2)
- g. เลือก Sample naming จะปรากฏหน้าต่าง Sample Naming (3)
- h. เลือก Include file name และ
- i. Include file extension
- j. เลือก OK

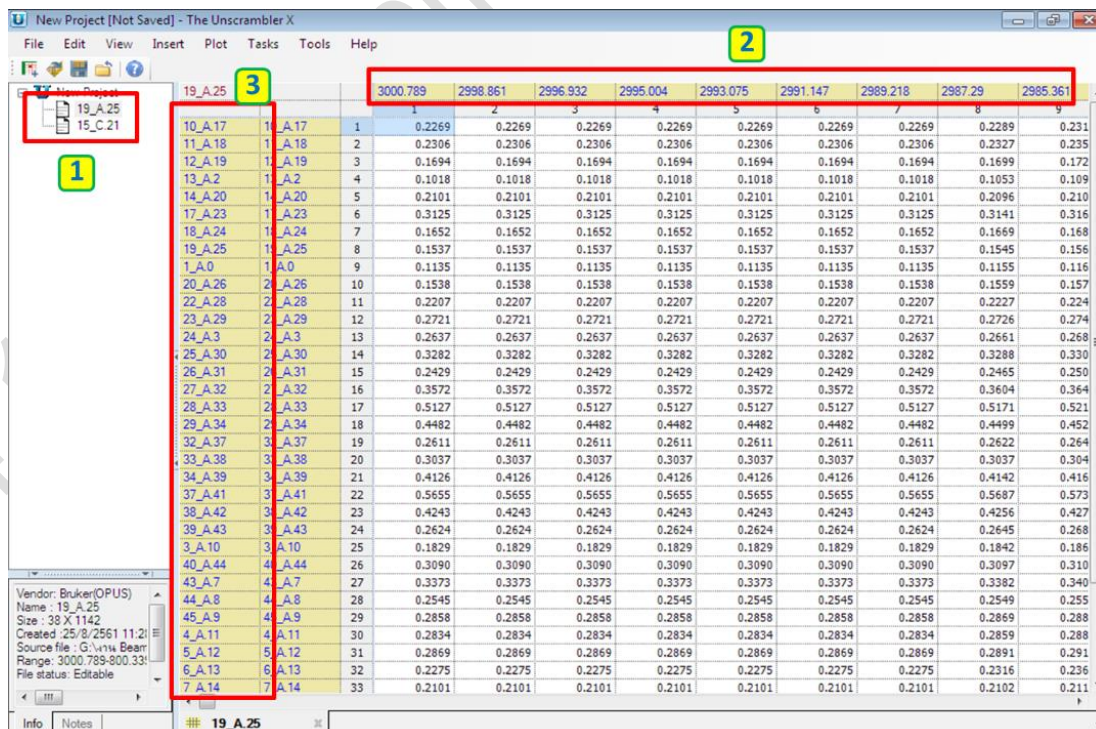


Import ข้อมูลกลุ่ม A และ C เข้ามา



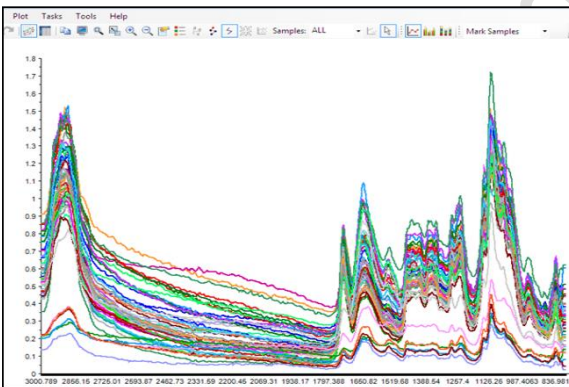
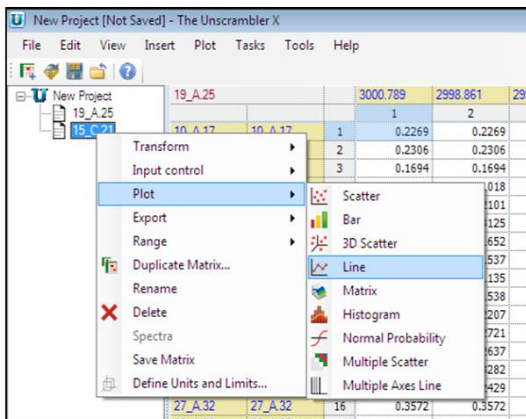
2. จะปรากฏข้อมูลสเปกตรัมที่ Import เข้ามา ประกอบด้วย

- กลุ่มของชุดข้อมูล A และ C หรือจะเรียกว่า data set หรือ data matrix (1)
- Column แสดงค่า Wavenumber (2)
- Row แสดงชื่อสเปกตรัมกลุ่มตัวอย่าง (3)

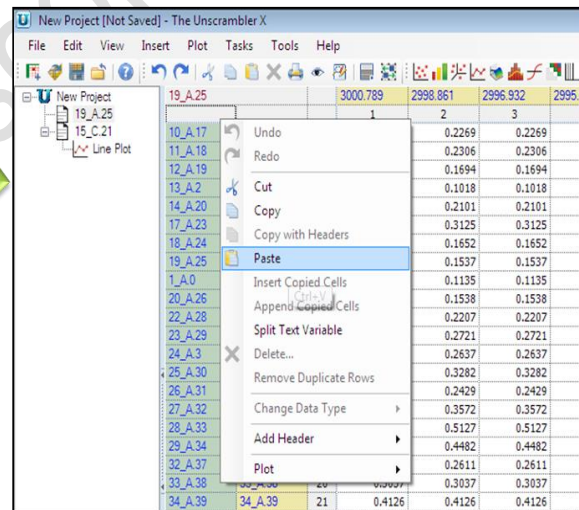
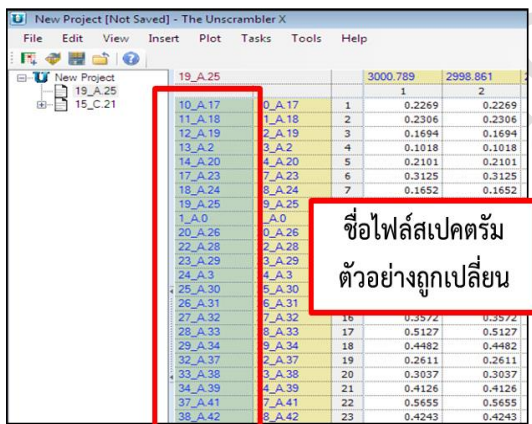
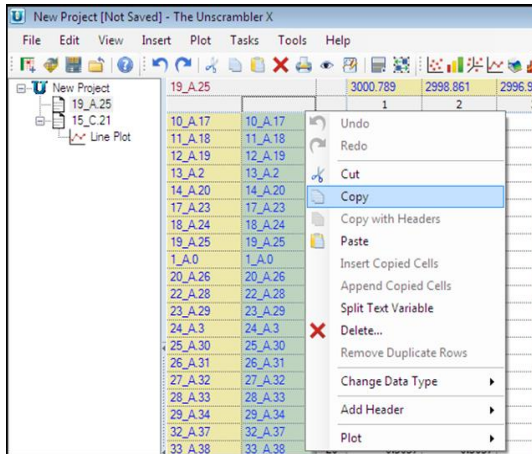


3. เมื่อ import data เข้ามาแล้วต้องการ plot graph เพื่อดูลักษณะสเปกตรัมของกลุ่มตัวอย่าง สามารถทำได้โดย

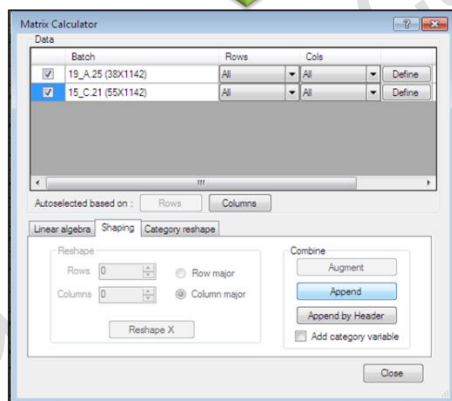
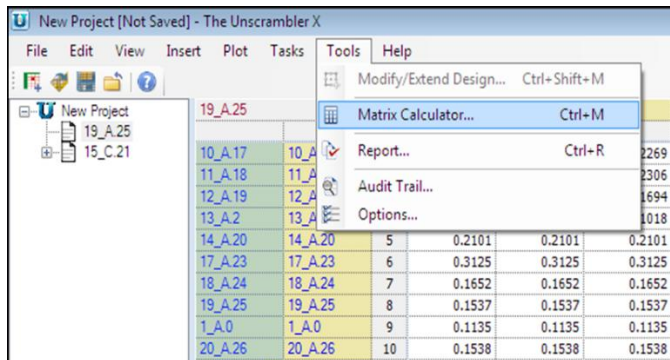
- คลิกขวา เลือก Plot
- เลือก Line จะปรากฏหน้าต่าง Line Plot-Scope
- กำหนด data matrix ที่ต้องการ Plot graph
- Plot type เลือก Sample
- เลือก OK จะปรากฏเส้น spectrum ของกลุ่มตัวอย่างที่เราเลือก plot



4. เนื่องจากชื่อไฟล์ spectrum จะแสดงใน column ที่ 2 เสมอ แต่เมื่อทำ data preprocess แล้ว column ที่ 2 จะหายไป ดังนั้นจึงควร copy column ที่ 2 มาวางใน column แรก และลบ column ที่ 2ทิ้งไปซึ่งทำได้โดย
- เลือก column ที่ 2
 - คลิกขวา เลือก copy
 - นำไปวางแทนที่ใน column ที่ 1

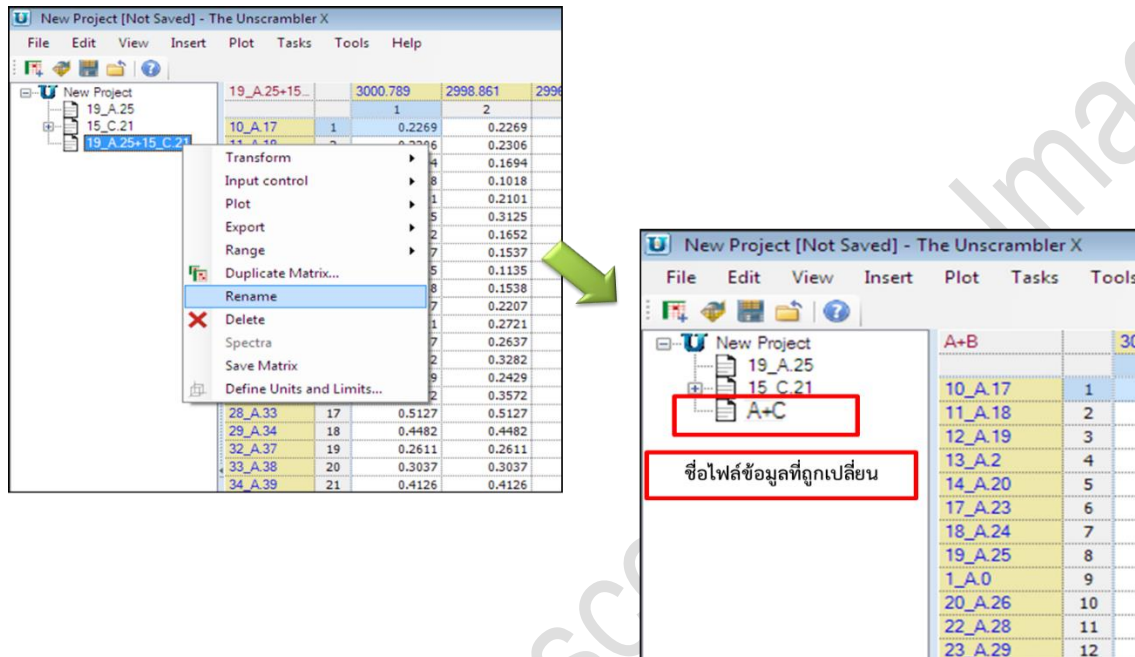


5. การวิเคราะห์ข้อมูลตัวอย่าง A และ C จะต้องทำการรวมข้อมูล spectrum ให้อยู่ใน data set เดียวกัน ซึ่งการรวมไฟล์ข้อมูล A และ C เข้าด้วยกัน ทำได้โดย
- เลือก Tool
 - เลือก Matrix Calculator จะปรากฏหน้าต่าง
 - Matrix Calculator
 - เลือกไฟล์ข้อมูลที่ต้องการรวม
 - เลือก Shaping
 - เลือก Append
 - เลือก OK จะปรากฏ file ข้อมูลที่นำมารวมกัน และจะแสดงข้อมูลชื่อ spectrum ของตัวอย่าง เฉพาะใน column แรกเท่านั้น

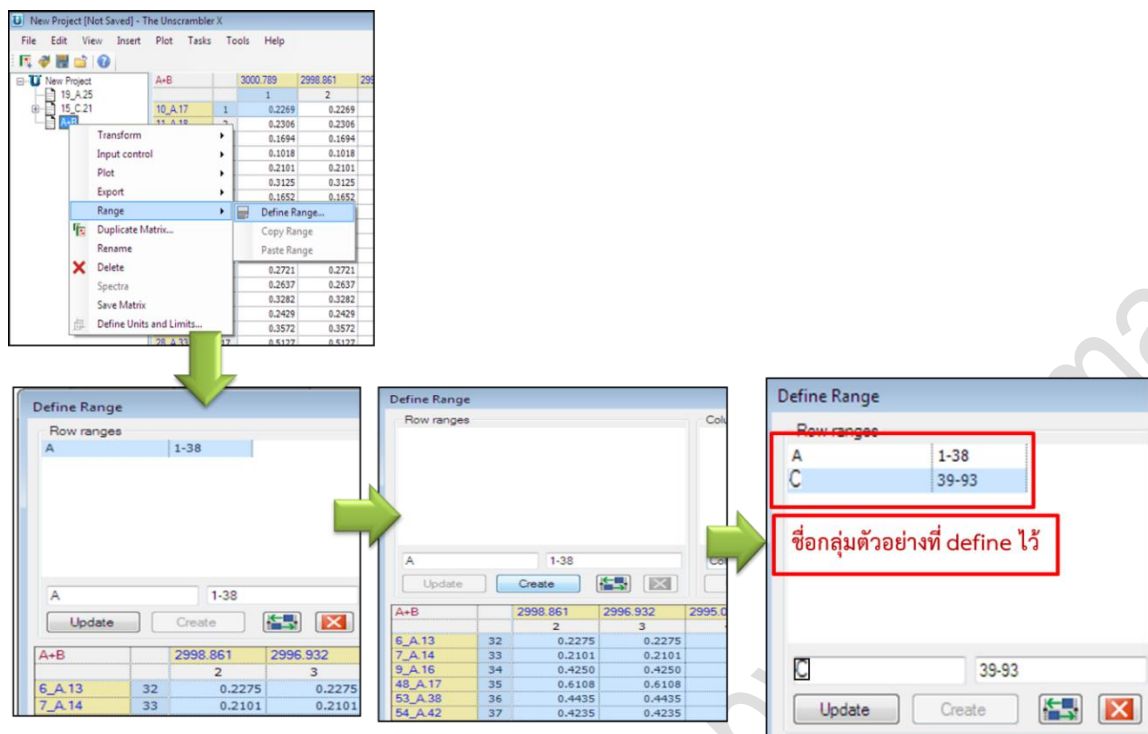


19_A_25+15_...	3000.789	2998.861	2
19_A_25	1	0.2269	0.2269
15_C_21	1	0.2306	0.2306
19_A_25+15_C_21	1	0.1694	0.1694
10_A_17	1	0.1018	0.1018
11_A_18	2	0.3125	0.3125
12_A_19	3	0.1652	0.1652
13_A_2	4	0.1537	0.1537
14_A_20	5	0.1135	0.1135
17_A_23	6	0.1538	0.1538
18_A_24	7	0.2207	0.2207
19_A_25	8		
1_A_0	9		
20_A_26	10		
22_A_28	11		

6. การเปลี่ยนชื่อไฟล์ชุดข้อมูล data matrix
 - a. คลิกขวาไฟล์ข้อมูลที่ต้องการเปลี่ยนชื่อ
 - b. เลือก Rename เพื่อทำการเปลี่ยนชื่อไฟล์



7. ในการวิเคราะห์ตัวอย่างนั้น ต้องมีการ define group ของตัวอย่างเพื่อให้ง่ายต่อการแสดงผล และการ define column number หรือ wavenumber เพื่อใช้ในการวิเคราะห์ตัวอย่าง ซึ่งสามารถทำได้โดย
8. การ Define Group ตัวอย่าง
 - a. คลิกขวา เลือก range ➡ define range หรือ
 - b. เลือก edit ➡ Define Range จะปรากฏหน้าต่าง Define Range
 - c. เลือกชุดตัวอย่างในแต่ละกลุ่มที่ต้องการ Define แล้วกำหนดชื่อ เช่น กลุ่มตัวอย่าง A และ C
 - d. กด Create จะปรากฏชื่อกลุ่มตัวอย่างด้านล่าง Row ranges
 - e. เลือก OK



9. จะปรากฏชื่อกลุ่มตัวอย่างที่ define ไว้ด้านล่าง data matrix ที่ define

New Project [Not Saved] - The Unscrambler X

File Edit View Insert Plot Tasks Tools Help

Rowset

A

C

Name: A+B

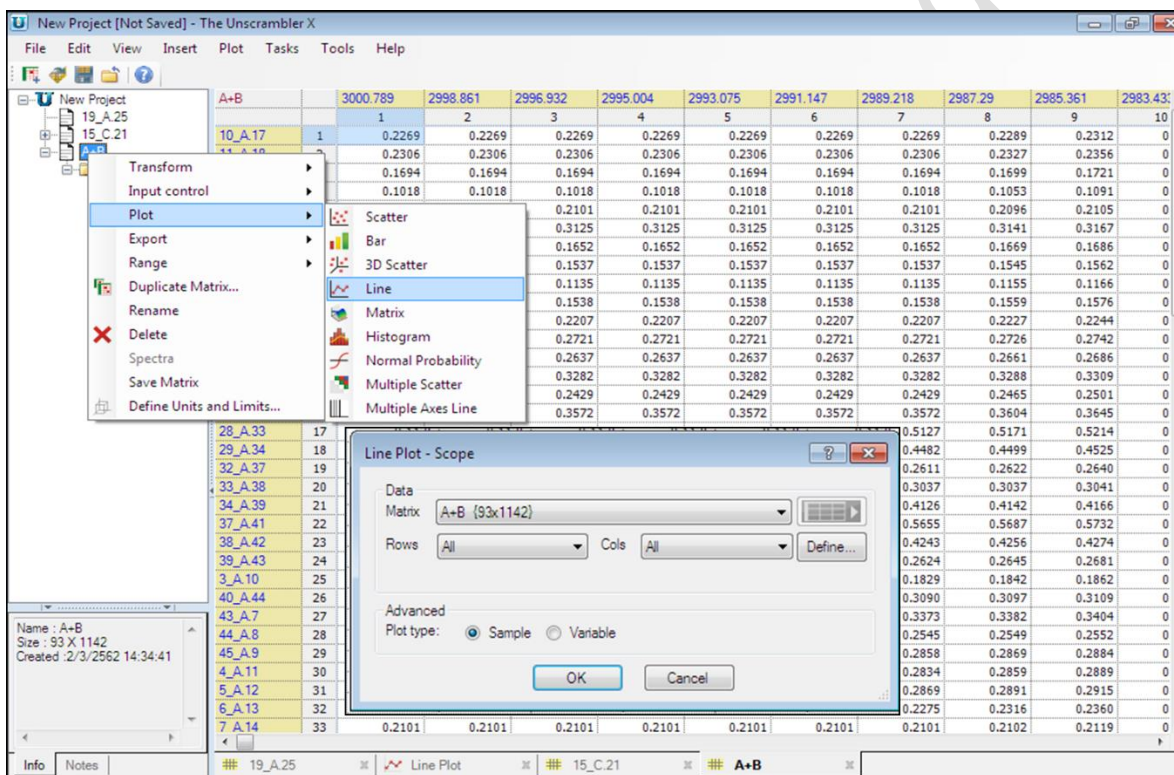
Size: 93 X 1142

Created: 2/3/2562 14:34:41

	A+B	3000.789	2998.861	2996.932	2995.004	2993.075	2991.147	2989.218	2987.29	2985.361	2983.432
10_A_17	1	0.2269	0.2269	0.2269	0.2269	0.2269	0.2269	0.2269	0.2269	0.2269	0.2269
11_A_18	2	0.2306	0.2306	0.2306	0.2306	0.2306	0.2306	0.2306	0.2306	0.2306	0.2306
12_A_19	3	0.1694	0.1694	0.1694	0.1694	0.1694	0.1694	0.1694	0.1694	0.1694	0.1694
13_A_20	4	0.1018	0.1018	0.1018	0.1018	0.1018	0.1018	0.1018	0.1018	0.1018	0.1018
14_A_21	5	0.2101	0.2101	0.2101	0.2101	0.2101	0.2101	0.2101	0.2101	0.2101	0.2101
15_A_22	6	0.3125	0.3125	0.3125	0.3125	0.3125	0.3125	0.3125	0.3125	0.3125	0.3125
16_A_23	7	0.1652	0.1652	0.1652	0.1652	0.1652	0.1652	0.1652	0.1652	0.1652	0.1652
17_A_24	8	0.1537	0.1537	0.1537	0.1537	0.1537	0.1537	0.1537	0.1537	0.1537	0.1537
18_A_25	9	0.1135	0.1135	0.1135	0.1135	0.1135	0.1135	0.1135	0.1135	0.1135	0.1135
19_A_26	10	0.1538	0.1538	0.1538	0.1538	0.1538	0.1538	0.1538	0.1538	0.1538	0.1538
20_A_27	11	0.2207	0.2207	0.2207	0.2207	0.2207	0.2207	0.2207	0.2207	0.2207	0.2207
21_A_28	12	0.2721	0.2721	0.2721	0.2721	0.2721	0.2721	0.2721	0.2721	0.2721	0.2721
22_A_29	13	0.2637	0.2637	0.2637	0.2637	0.2637	0.2637	0.2637	0.2637	0.2637	0.2637
23_A_30	14	0.3282	0.3282	0.3282	0.3282	0.3282	0.3282	0.3282	0.3282	0.3282	0.3282
24_A_31	15	0.2429	0.2429	0.2429	0.2429	0.2429	0.2429	0.2429	0.2429	0.2429	0.2429
25_A_32	16	0.3572	0.3572	0.3572	0.3572	0.3572	0.3572	0.3572	0.3572	0.3572	0.3572
26_A_33	17	0.5127	0.5127	0.5127	0.5127	0.5127	0.5127	0.5127	0.5127	0.5127	0.5127
27_A_34	18	0.4482	0.4482	0.4482	0.4482	0.4482	0.4482	0.4482	0.4482	0.4482	0.4482
28_A_35	19	0.2611	0.2611	0.2611	0.2611	0.2611	0.2611	0.2611	0.2611	0.2611	0.2611
29_A_36	20	0.3037	0.3037	0.3037	0.3037	0.3037	0.3037	0.3037	0.3037	0.3037	0.3037
30_A_37	21	0.4126	0.4126	0.4126	0.4126	0.4126	0.4126	0.4126	0.4126	0.4126	0.4126
31_A_38	22	0.5655	0.5655	0.5655	0.5655	0.5655	0.5655	0.5655	0.5655	0.5655	0.5655
32_A_39	23	0.4243	0.4243	0.4243	0.4243	0.4243	0.4243	0.4243	0.4243	0.4243	0.4243
33_A_40	24	0.2624	0.2624	0.2624	0.2624	0.2624	0.2624	0.2624	0.2624	0.2624	0.2624
34_A_41	25	0.1829	0.1829	0.1829	0.1829	0.1829	0.1829	0.1829	0.1829	0.1829	0.1829
35_A_42	26	0.3090	0.3090	0.3090	0.3090	0.3090	0.3090	0.3090	0.3090	0.3090	0.3090
36_A_43	27	0.3373	0.3373	0.3373	0.3373	0.3373	0.3373	0.3373	0.3373	0.3373	0.3373
37_A_44	28	0.2545	0.2545	0.2545	0.2545	0.2545	0.2545	0.2545	0.2545	0.2545	0.2545
38_A_45	29	0.2858	0.2858	0.2858	0.2858	0.2858	0.2858	0.2858	0.2858	0.2858	0.2858
39_A_46	30	0.2834	0.2834	0.2834	0.2834	0.2834	0.2834	0.2834	0.2834	0.2834	0.2834
40_A_47	31	0.2869	0.2869	0.2869	0.2869	0.2869	0.2869	0.2869	0.2869	0.2869	0.2869
41_A_48	32	0.2275	0.2275	0.2275	0.2275	0.2275	0.2275	0.2275	0.2275	0.2275	0.2275
42_A_49	33	0.2101	0.2101	0.2101	0.2101	0.2101	0.2101	0.2101	0.2101	0.2101	0.2101

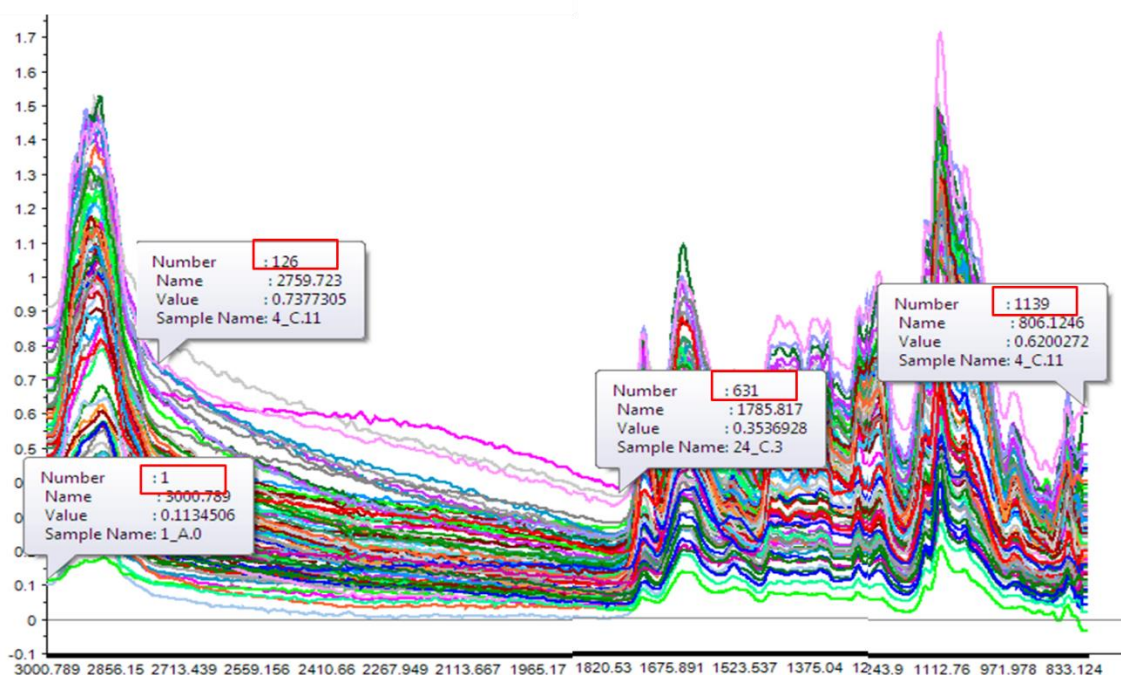
10. การ define column นั้น ก่อนที่จะทำการ define column เพื่อใช้ในการวิเคราะห์ข้อมูล จะต้องทราบช่วงของ wavenumber ที่ต้องการใช้ในการวิเคราะห์ก่อน โดยสามารถทราบช่วง wavenumber ได้จากการ plot graph และจดค่าตัวเลข column ที่สัมพันธ์กับ wavenumber

- คลิกขวา เลือก Plot
- เลือก Line จะปรากฏหน้าต่าง Line Plot-Scope
- กำหนด matrix ที่ต้องการ Plot
- เลือก Sample
- เลือก OK



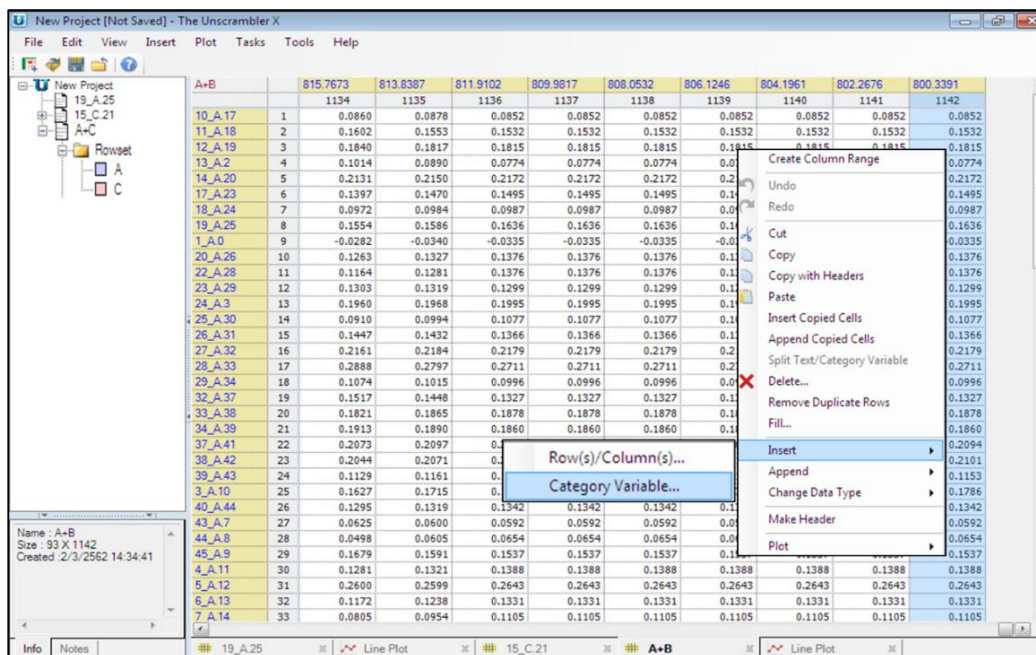
11. การ define column จะต้องมีการกำหนดช่วงที่ต้องการ define คือช่วง wavenumber ที่ต้องการนำไปใช้ในการวิเคราะห์

- a. การกำหนดช่วง column ต้องใส่ column number ที่อยู่ในช่วง wavenumber ที่เราต้องการใช้
- b. ในตัวอย่างให้กำหนดช่วง column number 2 ช่วงคือ
 - i. ที่ 1-126 (wn. 3000-2759 cm^{-1}) และ 631-1139 (wn. 1785-806 cm^{-1}) กำหนดให้เป็นช่วง Use คือช่วงที่ต้องการนำมาใช้ในการเปรียบเทียบ
 - ii. ที่ 1-1141 (wn. 3000-802 cm^{-1}) กำหนดให้เป็นช่วง Specific range คือช่วงทั้งหมดของ spectrum

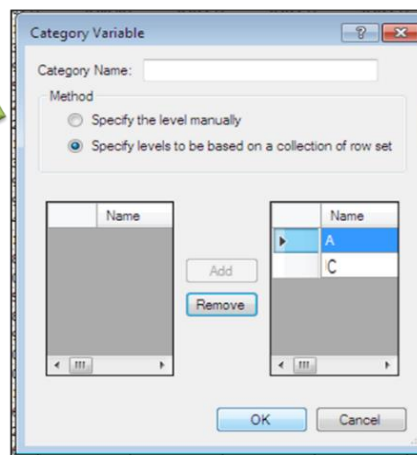
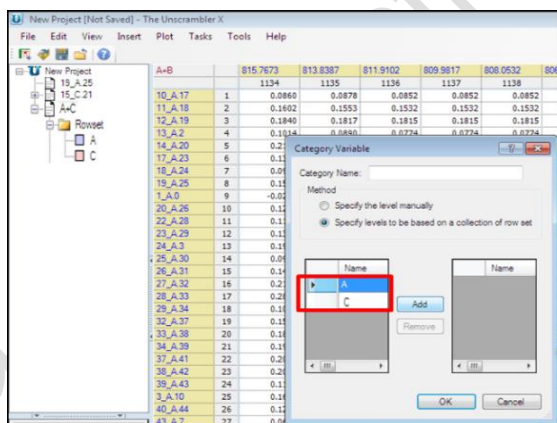


12. Define column เพื่อกำหนดกลุ่มสเปกตรัมตัวอย่าง A และ C

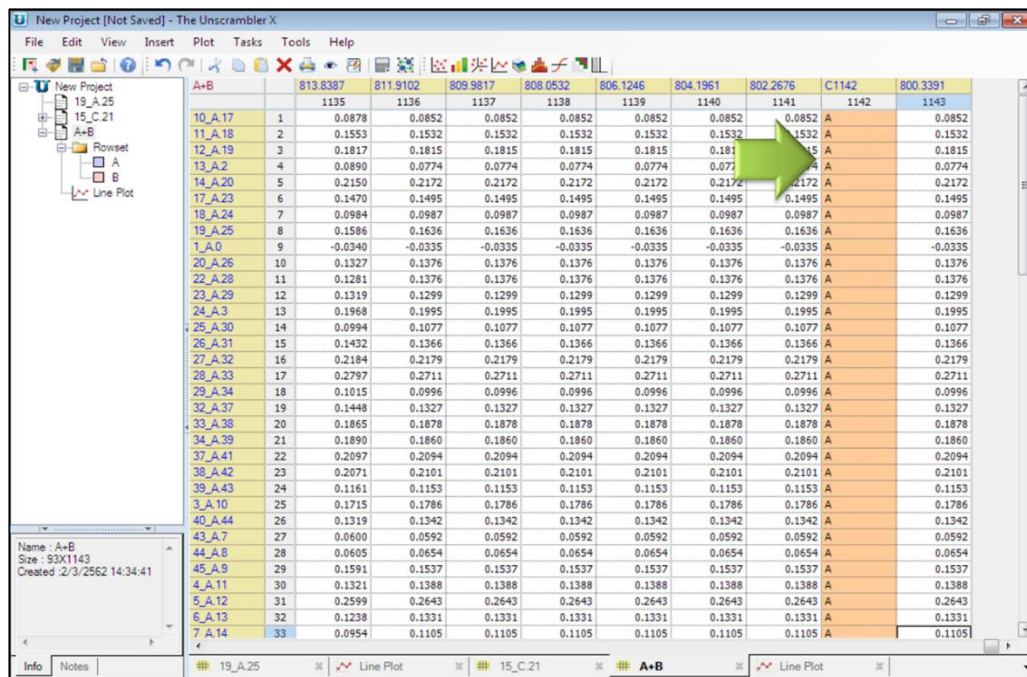
- a. เลือก column สุดท้าย
- b. คลิกขวา
- c. เลือก Insert
- d. เลือก Category Variable



- จะปรากฏหน้าต่าง Category Name
- เลือก Specify levels to be based on a collection of row set
- เลือก Add ตัวอย่างทุกกลุ่มไปช่องด้านขวามือ
- เลือก OK

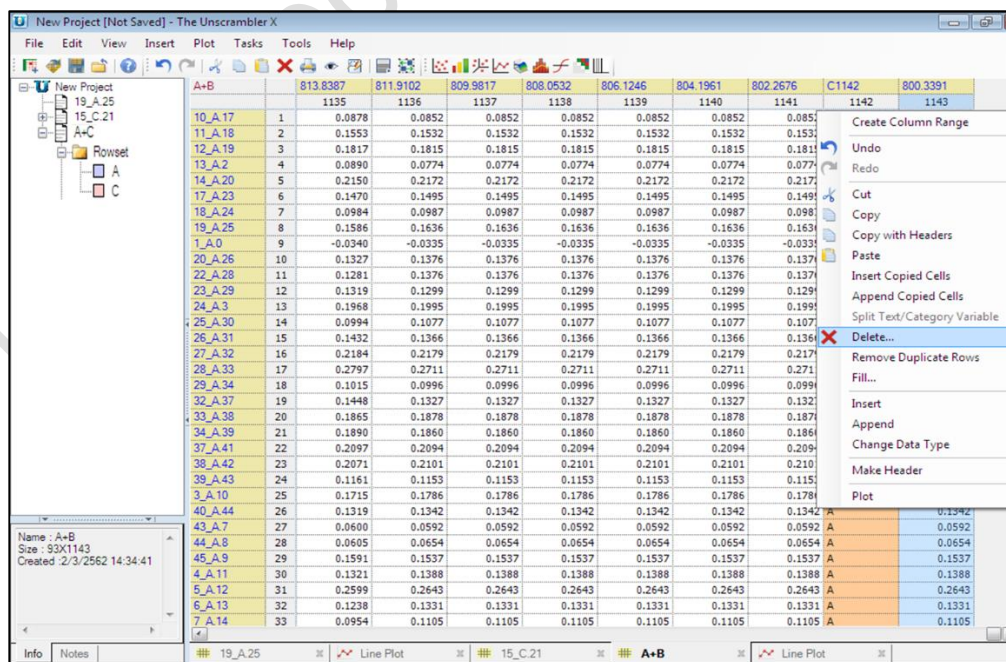


13. จะปรากฏ column ของตัวอย่างแต่ละกลุ่มเพิ่มขึ้น



The screenshot shows the 'New Project [Not Saved] - The Unscrambler X' window. The main data table has columns for variables (A+B, 813.8387, 811.9102, 809.9817, 808.0532, 806.1246, 804.1961, 802.2676, C1142, 800.3391) and rows for samples (10_A.17, 11_A.18, 12_A.19, 13_A.2, 14_A.20, 17_A.23, 18_A.24, 19_A.25, 1_A.0, 20_A.26, 22_A.28, 23_A.29, 24_A.3, 25_A.30, 26_A.31, 27_A.32, 28_A.33, 29_A.34, 32_A.37, 33_A.38, 34_A.39, 37_A.41, 38_A.42, 39_A.43, 3_A.10, 40_A.44, 43_A.7, 44_A.8, 45_A.9, 4_A.11, 5_A.12, 6_A.13, 7_A.14). A green arrow points to the 'C1142' column header.

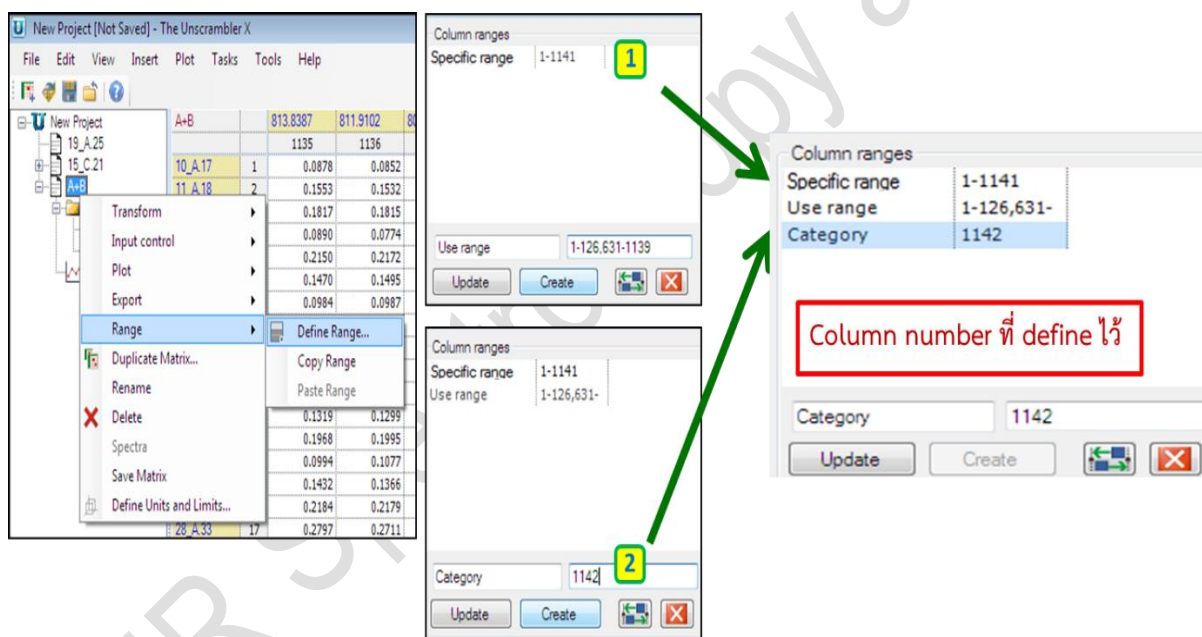
- ลบ column สุดท้ายออก
- คลิกขวาใน column สุดท้าย
- เลือก delete



The screenshot shows the same data table as above, but with a context menu open over the 'C1142' column header. The menu options include: Create Column Range, Undo, Redo, Cut, Copy, Copy with Headers, Paste, Insert Copied Cells, Append Copied Cells, Split Text/Category Variable, Delete... (highlighted with a red X), Remove Duplicate Rows, Fill..., Insert, Append, Change Data Type, Make Header, and Plot.

14. การกำหนดช่วง wavenumber ในการ define column

- เลือก Edit
- เลือก Define Range จะปรากฏหน้าต่าง Define Range
- กำหนดช่วง Specific range
- (ช่วงทั้งหมดยกเว้น Category) กด Create
- ใส่หมายเลข column (ช่วง wavenumber) ที่ใช้ในการวิเคราะห์ แล้วกำหนดชื่อ use range กด Create
- ใส่หมายเลข column สุดท้าย กำหนด Category ของชื่อ กลุ่มตัวอย่าง A และ C แล้วกด Create
- เลือก OK



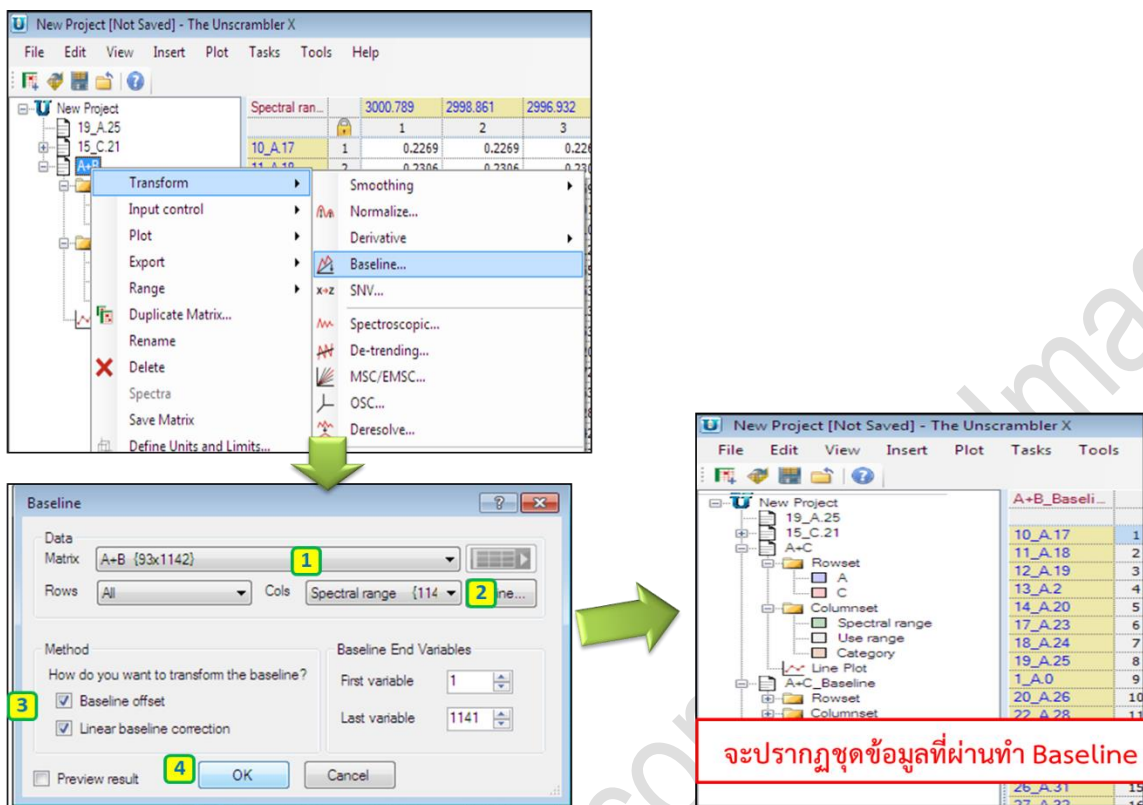
15. จะปรากฏ Range คือ group ตัวอย่างและช่วง Column ของ wavenumber ที่ define ได้ข้อมูล

Spectral ran...	3000.789	2998.861	2996.932	2995.004	2993.075	2991.147	2989.218	2987.29	2985.361
10_A.17	1	0.2269	0.2269	0.2269	0.2269	0.2269	0.2269	0.2269	0.2312
11_A.18	2	0.2306	0.2306	0.2306	0.2306	0.2306	0.2306	0.2306	0.2356
12_A.19	3	0.1694	0.1694	0.1694	0.1694	0.1694	0.1694	0.1694	0.1721
13_A.2	4	0.1018	0.1018	0.1018	0.1018	0.1018	0.1018	0.1018	0.1091
14_A.20	5	0.2101	0.2101	0.2101	0.2101	0.2101	0.2101	0.2101	0.2105
17_A.23	6	0.3125	0.3125	0.3125	0.3125	0.3125	0.3125	0.3125	0.3167
18_A.24	7	0.1652	0.1652	0.1652	0.1652	0.1652	0.1652	0.1652	0.1686
19_A.25	8	0.1537	0.1537	0.1537	0.1537	0.1537	0.1537	0.1537	0.1562
1_A.0	9	0.1135	0.1135	0.1135	0.1135	0.1135	0.1135	0.1135	0.1166
20_A.26	10	0.1538	0.1538	0.1538	0.1538	0.1538	0.1538	0.1538	0.1576
22_A.28	11	0.2207	0.2207	0.2207	0.2207	0.2207	0.2207	0.2207	0.2244
23_A.29	12	0.2721	0.2721	0.2721	0.2721	0.2721	0.2721	0.2721	0.2742
24_A.3	13	0.2637	0.2637	0.2637	0.2637	0.2637	0.2637	0.2637	0.2686
25_A.30	14	0.3282	0.3282	0.3282	0.3282	0.3282	0.3282	0.3282	0.3309
26_A.31	15	0.2429	0.2429	0.2429	0.2429	0.2429	0.2429	0.2429	0.2501
27_A.32	16	0.3572	0.3572	0.3572	0.3572	0.3572	0.3572	0.3572	0.3645
28_A.33	17	0.5127	0.5127	0.5127	0.5127	0.5127	0.5127	0.5127	0.5214
29_A.34	18	0.4482	0.4482	0.4482	0.4482	0.4482	0.4482	0.4482	0.4525
32_A.37	19	0.2611	0.2611	0.2611	0.2611	0.2611	0.2611	0.2611	0.2640
33_A.38	20	0.3037	0.3037	0.3037	0.3037	0.3037	0.3037	0.3037	0.3041
34_A.39	21	0.4126	0.4126	0.4126	0.4126	0.4126	0.4126	0.4126	0.4166
37_A.41	22	0.5655	0.5655	0.5655	0.5655	0.5655	0.5655	0.5655	0.5732
38_A.42	23	0.4243	0.4243	0.4243	0.4243	0.4243	0.4243	0.4243	0.4274
39_A.43	24	0.2624	0.2624	0.2624	0.2624	0.2624	0.2624	0.2624	0.2681
3_A.10	25	0.1829	0.1829	0.1829	0.1829	0.1829	0.1829	0.1829	0.1862
40_A.44	26	0.3090	0.3090	0.3090	0.3090	0.3090	0.3090	0.3090	0.3109
43_A.7	27	0.3373	0.3373	0.3373	0.3373	0.3373	0.3373	0.3373	0.3404
44_A.8	28	0.2545	0.2545	0.2545	0.2545	0.2545	0.2545	0.2545	0.2552
45_A.9	29	0.2858	0.2858	0.2858	0.2858	0.2858	0.2858	0.2858	0.2884
4_A.11	30	0.2834	0.2834	0.2834	0.2834	0.2834	0.2834	0.2834	0.2889
5_A.12	31	0.2869	0.2869	0.2869	0.2869	0.2869	0.2869	0.2869	0.2915
6_A.13	32	0.2275	0.2275	0.2275	0.2275	0.2275	0.2275	0.2275	0.2316
7_A.14	33	0.2101	0.2101	0.2101	0.2101	0.2101	0.2101	0.2101	0.2119

16. การทำ data-preprocessing

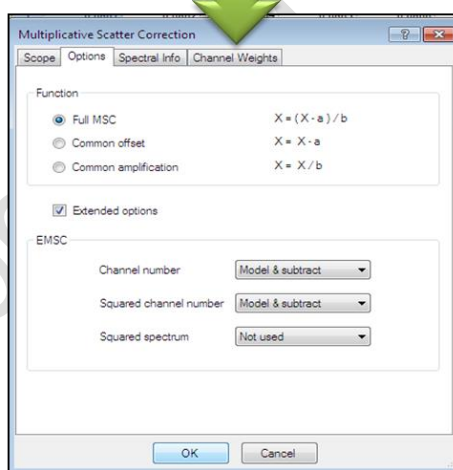
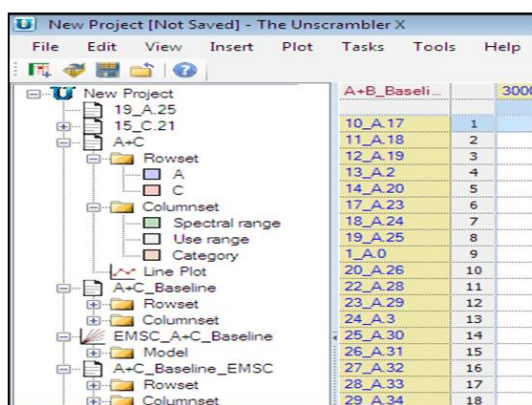
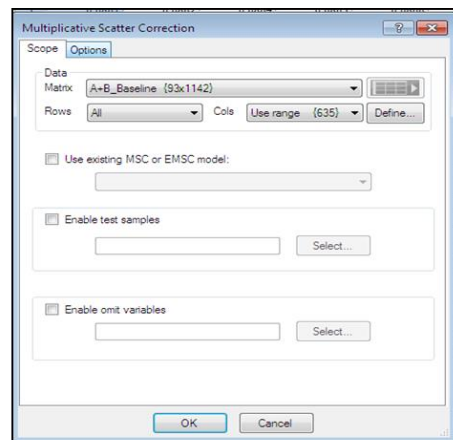
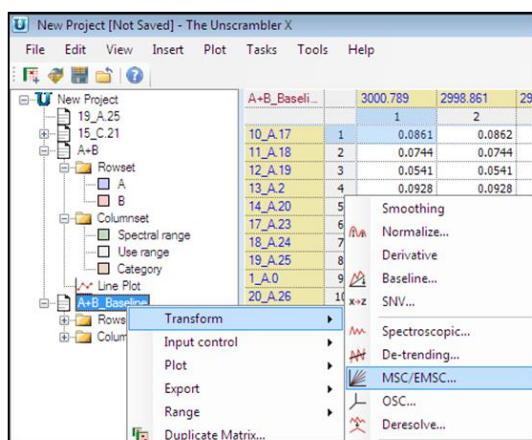
Data preprocess ในตัวอย่างนี้ จะประกอบด้วย: Baseline, MSC/EMSC, PCA โดยจะต้องทำเป็นลำดับต่อกันไป โดยเริ่มจากการทำ Baseline เพื่อปรับให้ฐานพีคมีความเรียบเสมอกัน

- คลิกขวา เลือก Transform
- เลือก Baseline
- จะปรากฏหน้าต่างให้กำหนดค่า Parameters
- เลือก Data Matrix ที่ต้องการทำ Baseline (1)
- เลือก column ที่ define ไว้ (2)
- เลือก Baseline offset และ (3)
- Linear baseline correction (3)
- กด Preview result เพื่อให้เห็นแสดงผล Baseline
- เลือก OK (4)



17. การทำ MSC/EMSC เป็นการทำให้ spectral normalization

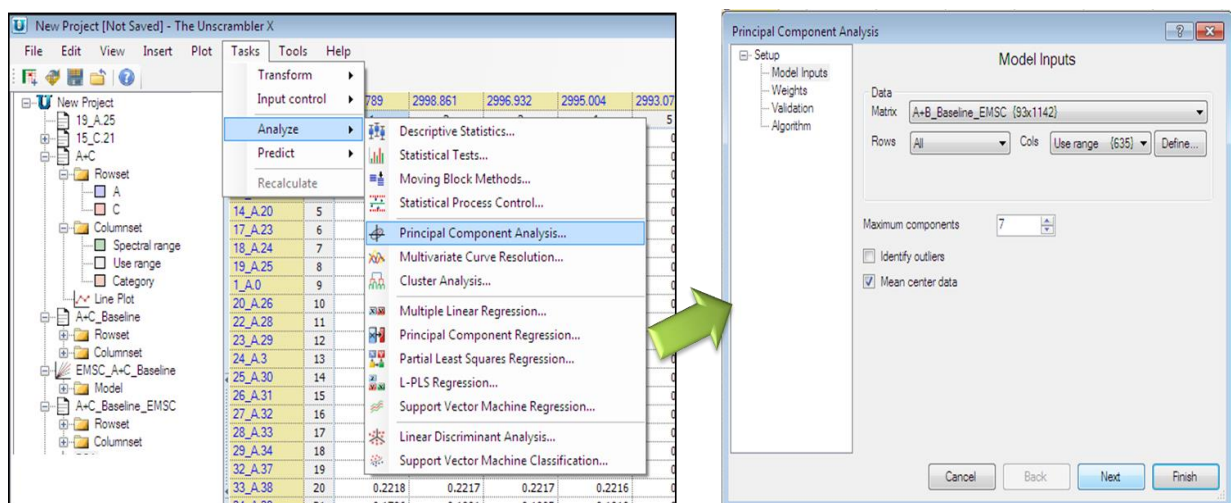
- คลิกขวาเลือก Transform
- เลือก MSC/EMSC
- จะปรากฏหน้าต่าง Multiplicative Scatter Correction
- ที่หน้า Scope เลือก Data Matrix ที่ต้องการทำ EMSC
- เลือก Column (Use) ที่ define ไว้
- ที่หน้า Options เลือก Extended options
- เลือก OK



จะปรากฏชุดข้อมูลที่ทำผ่านทำ EMSC

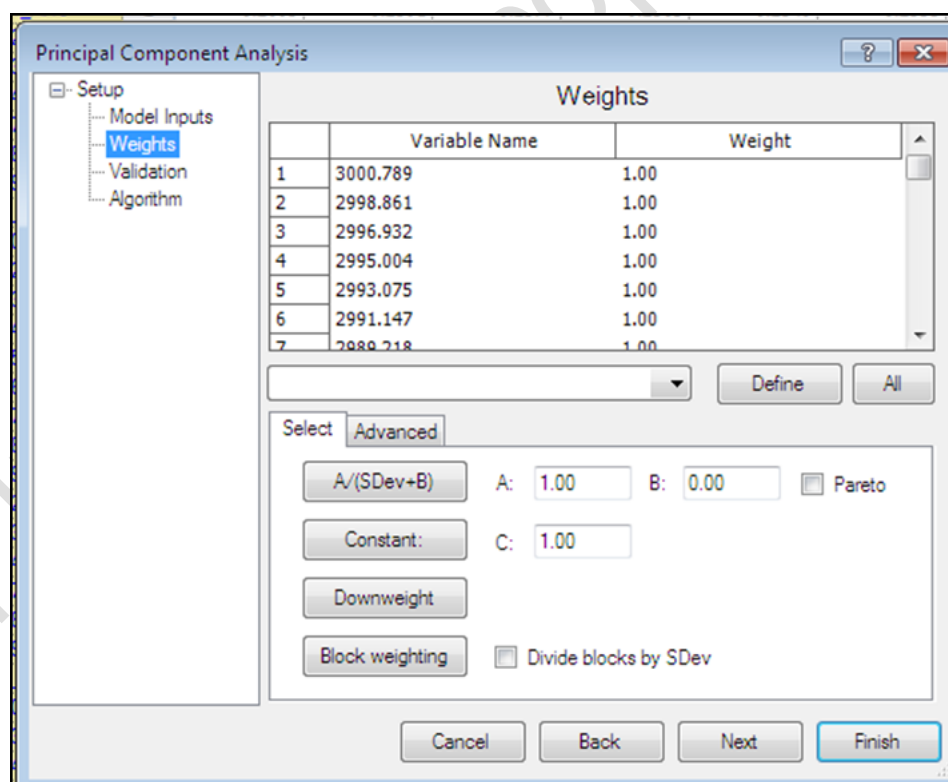
18. การทำ Principal component analysis (PCA) เพื่อแยกความแตกต่างของตัวอย่าง A และ C

- เลือก Tasks
- เลือก Analyze
- เลือก Principal Component Analysis
- จะปรากฏหน้าต่าง Principal Component Analysis ให้กำหนดค่า Setup
- กำหนดค่า Model Inputs
- เลือก Data Matrix ที่
- ต้องการทำ PCA และเลือก Column ที่ define ไว้
- เลือก Next



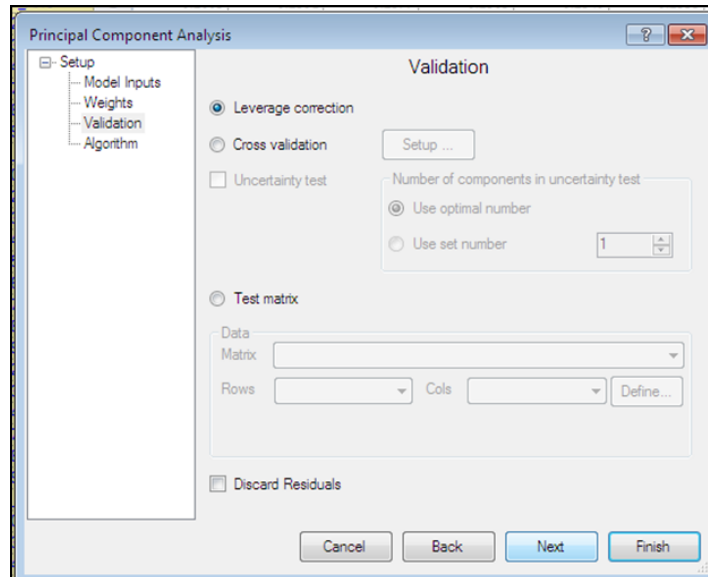
19. จะปรากฏหน้าต่างให้กำหนดค่า weights

- กำหนดค่า Weights
- Next



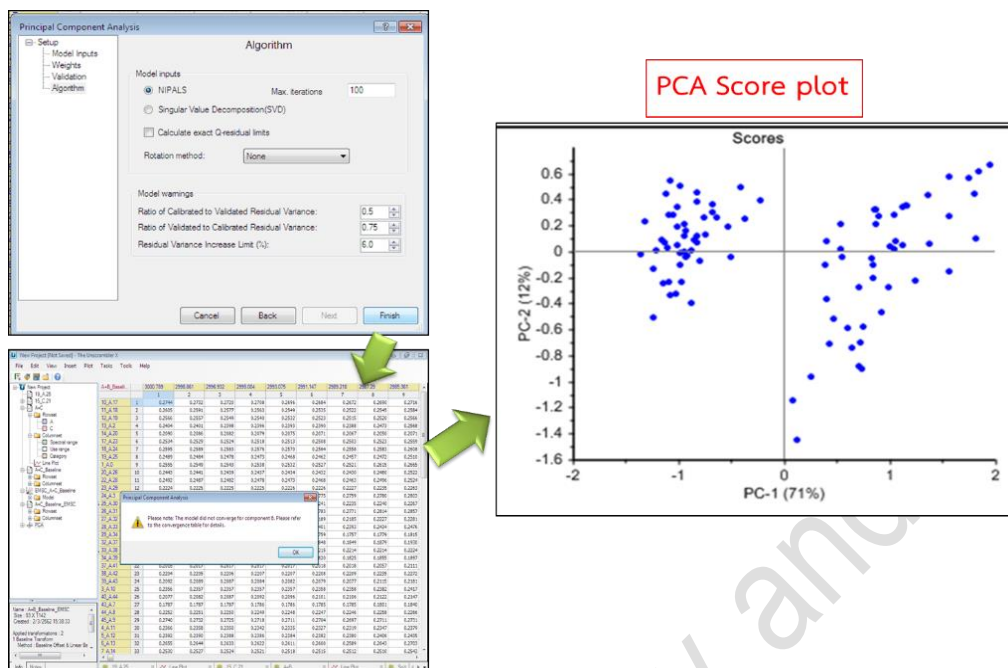
20. จะปรากฏหน้าต่างให้กำหนดค่า validation

- กำหนดค่า Validation
- เลือก Leverage correction
- Next



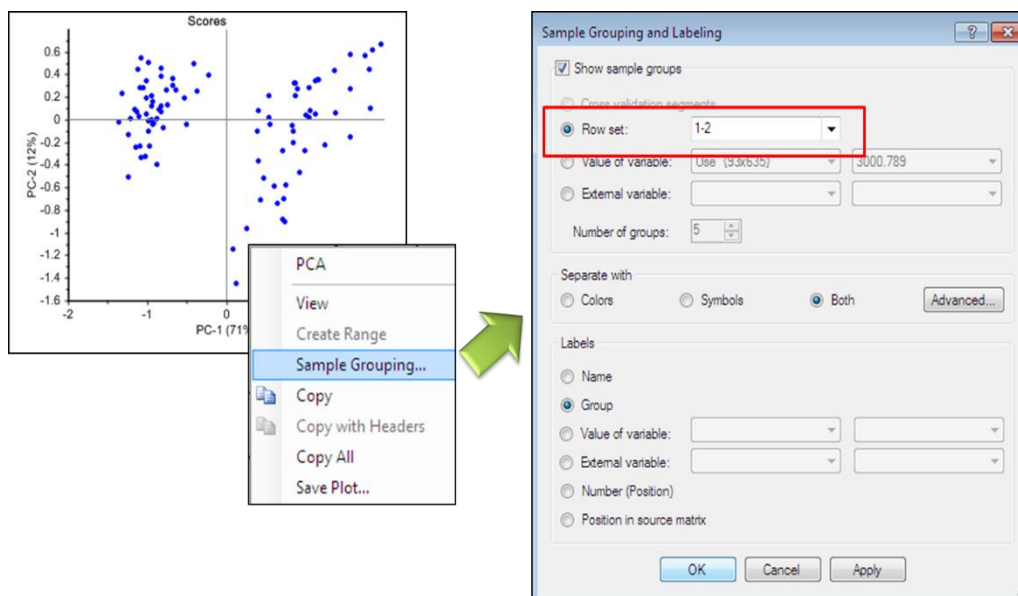
21. จะปรากฏหน้าต่างให้กำหนดค่า Algorithm

- กำหนดค่า Algorithm
- Model inputs เลือก NIPALS
- Finish
- กด OK
- จะปรากฏหน้าต่างแสดงผลการทำ PCA



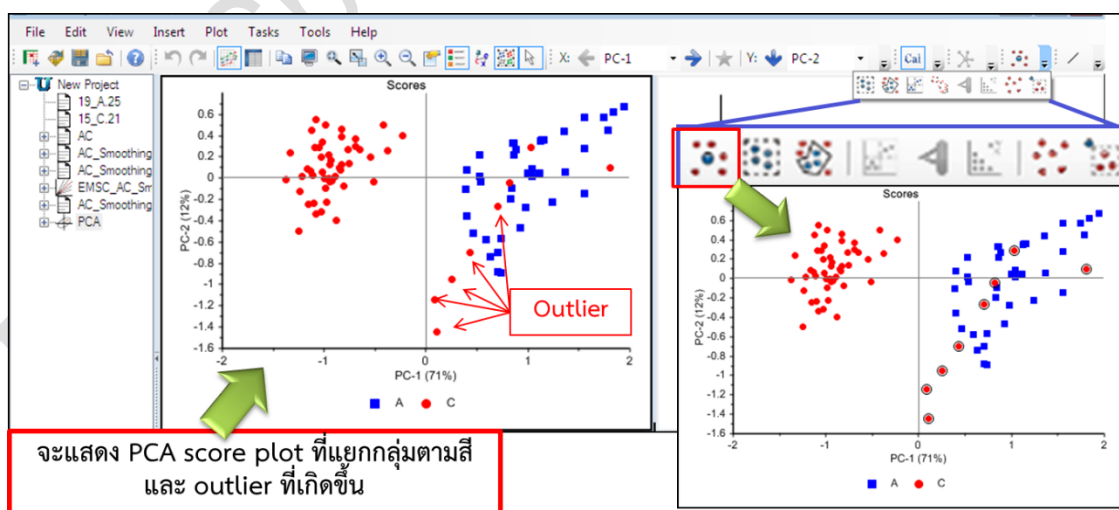
22. จากกราฟ PCA score plot ที่ได้ ถ้าต้องการให้แสดงสีเพื่อแยกกลุ่มตัวอย่างออก สามารถทำได้โดย

- คลิกขวาวบริเวณหน้าต่างแสดง PCA score plot
- เลือก Sample Grouping...
- จะปรากฏหน้าต่าง Sample Grouping and Labeling
- เลือก Show sample groups
- เลือก Row set:
- Separate with (เลือก Color หรือ Both)
- Labels (เลือก Group)
- กด OK



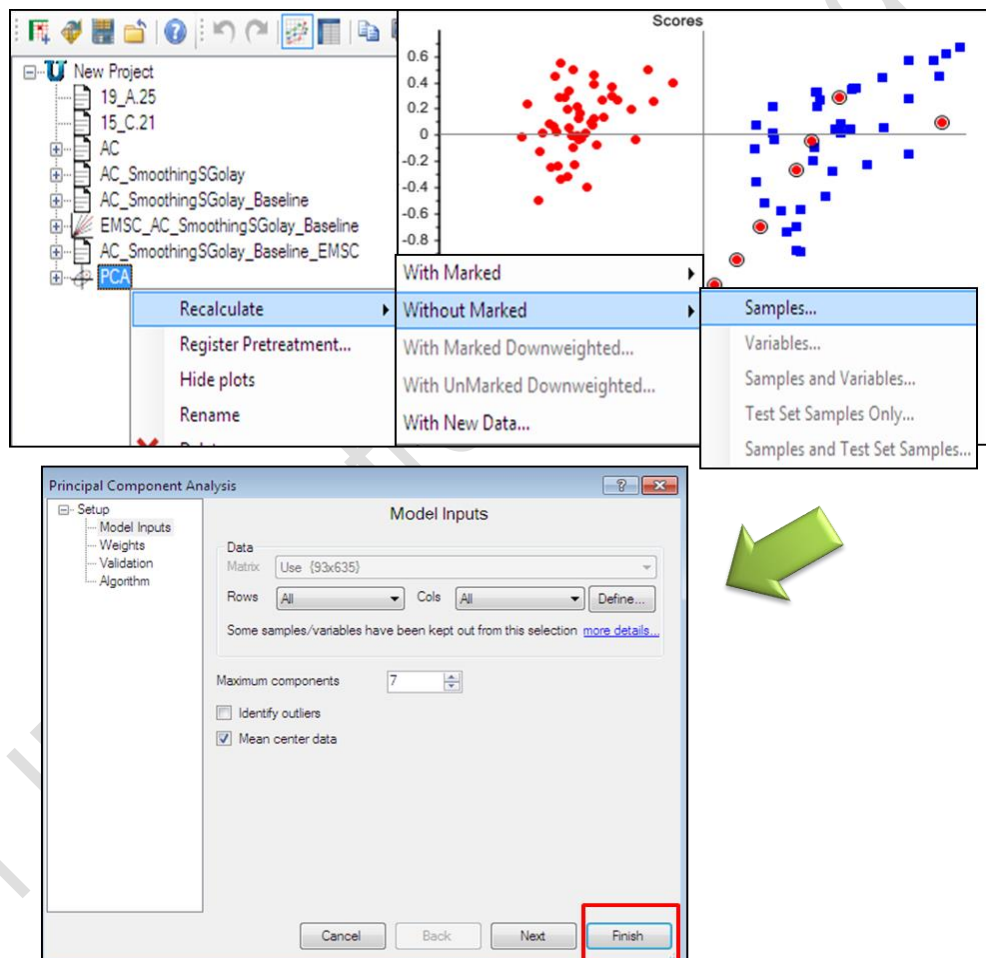
23. จะปรากฏกราฟ PCA score plot ที่แยกสีตามกลุ่มตัวอย่าง โดยจากกราฟจะแสดงให้เห็นว่า ในกลุ่มตัวอย่าง A มี outlier ที่เกิดจากตัวอย่าง B ปนอยู่ ตัวอย่าง โดยการตัด outlier ออก แล้วทำกราฟ PCA ใหม่ นั้น จะต้องมีการตัด spectrum ที่เป็น outlier ออก ซึ่งสามารถทำได้โดย

- เลือก mark จุดที่เป็น outlier โดยเลือกจาก icon ด้านบนขวามือ ประกอบด้วย
- Mark one by one 
- หรือ Mark with rectangular 
- (เมื่อต้องการลบจุดที่ mark ให้คลิกซ้ำที่จุดเดิมหรือเลือก Unmark all )

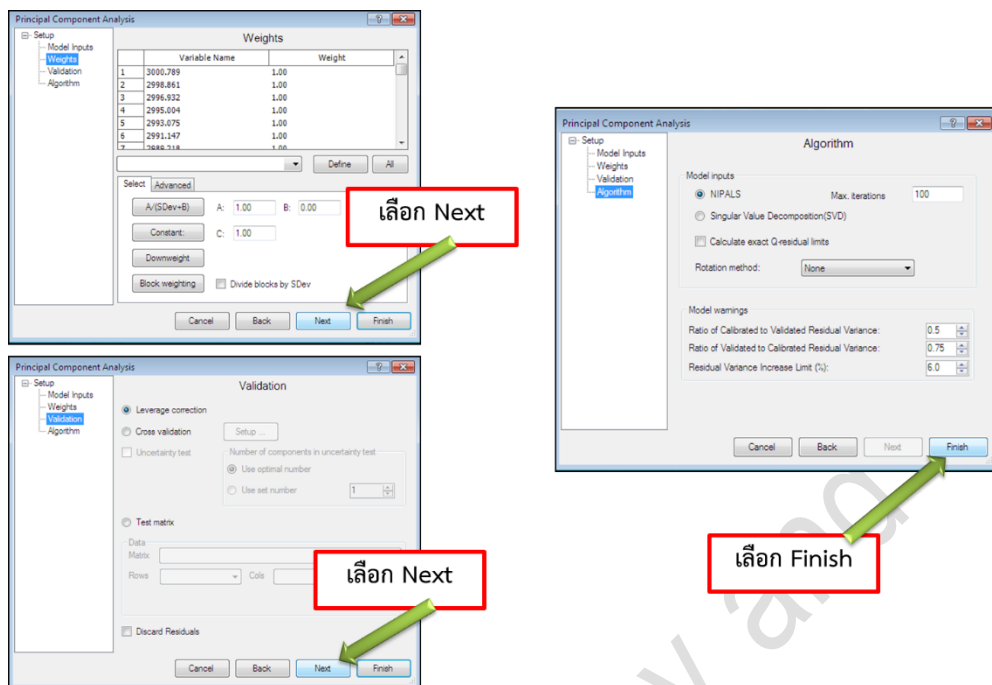


24. การตัด outlier ออกจากกลุ่มตัวอย่าง แล้ว plot กราฟ PCA score plot ขึ้นใหม่หลัง สามารถทำได้โดย

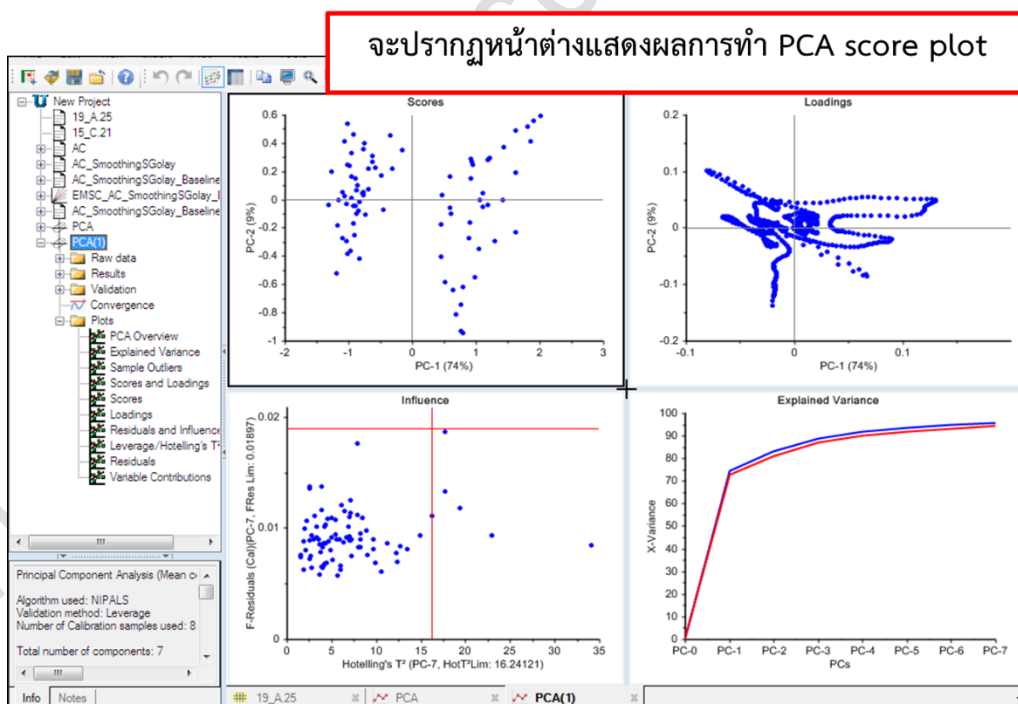
- mark จุดที่ต้องการตัดออกแล้ว
- คลิกขวาวบริเวณ PCA
- เลือก Recalculate
- เลือก Without Marked
- เลือก Sample...
- จะปรากฏหน้าต่างให้ทำ Principal Component Analysis
- เลือก Next



- หน้าต่างให้กำหนดค่า weights เลือก next
- หน้าต่างให้กำหนดค่า validation เลือก next
- หน้าต่างให้กำหนดค่า Algorithm เลือก finish

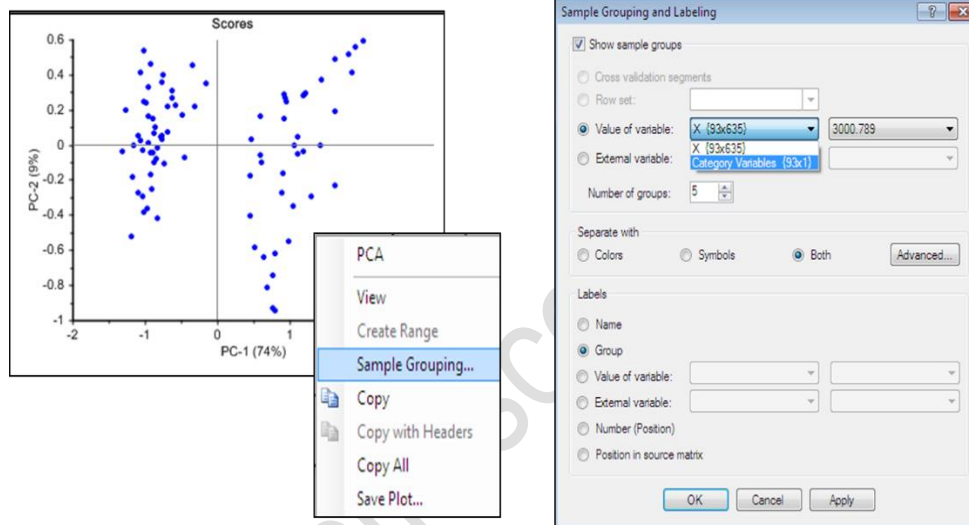


25. จะปรากฏหน้าต่างแสดงผลการทำ PCA

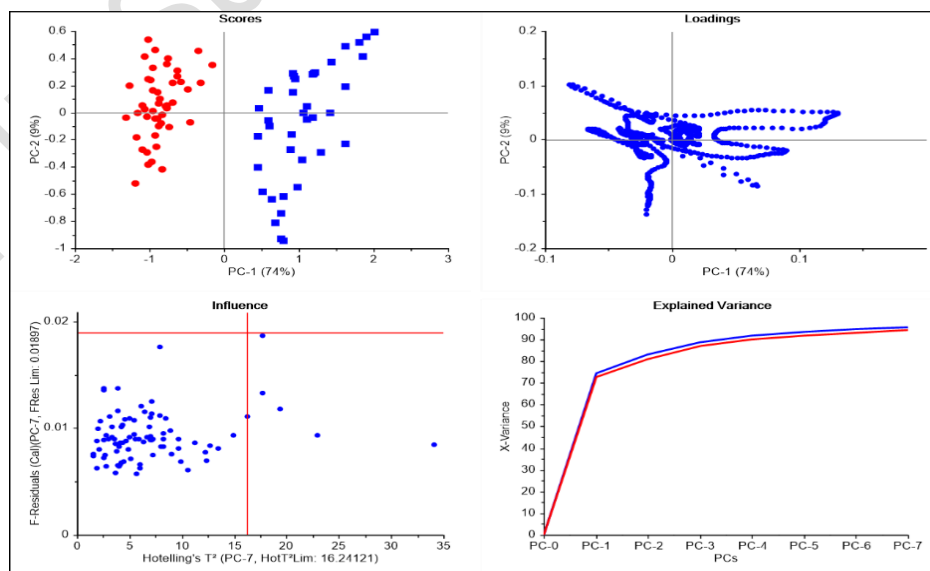


26. เมื่อต้องการให้กราฟ PCA score plot แสดงสีแยกตามตัวอย่าง สามารถทำได้โดย

- คลิกขวาวบริเวณหน้าต่างแสดง PCA score plot
- เลือก Sample Grouping...
- จะปรากฏหน้าต่าง Sample Grouping and Labeling
- เลือก Show sample groups
- เลือก Value of variable: Category Variables
- กด OK



27. จะปรากฏหน้าต่างแสดงผลการทำ PCA score plot ที่ผ่านการตัด Outlier ออกแล้ว

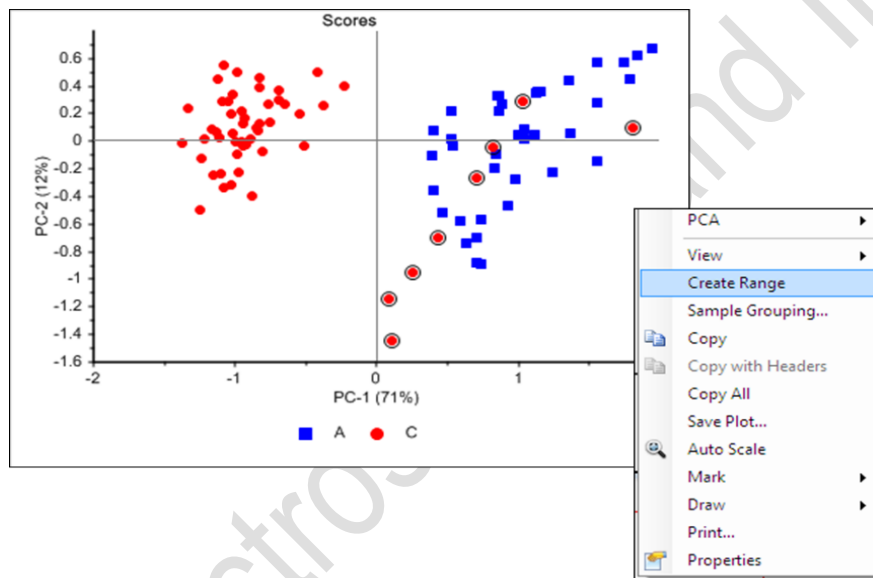


28. การแสดงผลการตัด outlier ประกอบด้วยการแสดงผลในชุดข้อมูลดังนี้

- แสดงข้อมูลชุดที่ไม่มี outlier (Training)
- แสดงข้อมูลสเปกตรัมที่เป็น outlier

โดยสามารถเลือกให้แสดงผล spectrum ที่เป็น outlier หรือให้แสดงผลเป็น data matrix ของชุด data set ใหม่ได้ โดยหลังจากที่ mark จุดที่เป็น outlier แล้ว

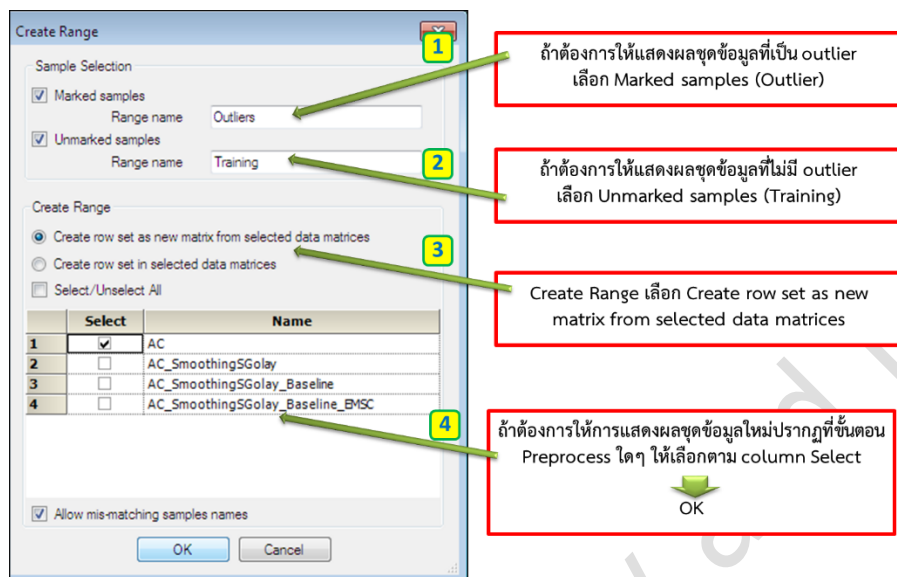
- คลิกขวา เลือก Create Range
- จะปรากฏหน้าต่าง Create Range



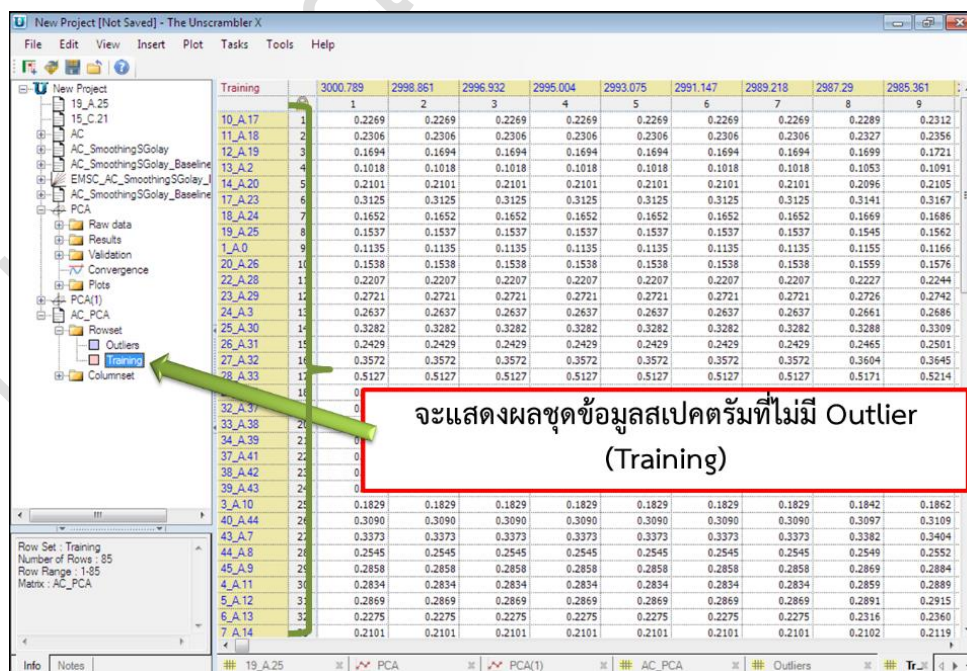
29. โดยสามารถตั้งค่าการแสดงผลได้ดังนี้

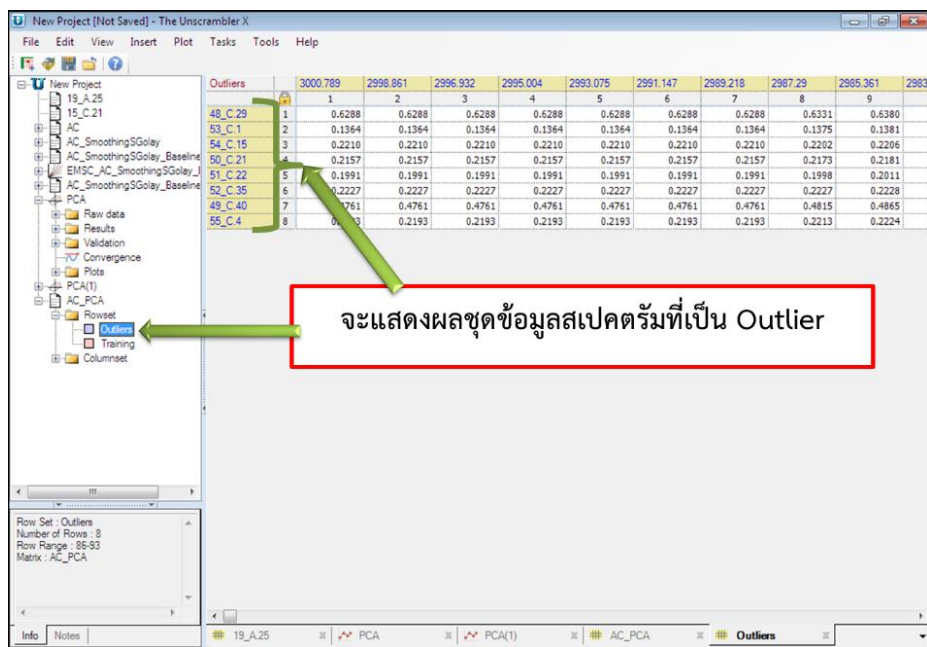
- 1) ใน sample selection จะเป็นการแสดงผลชุดข้อมูล data set ใหม่ โดยสามารถกำหนดให้แสดงผล data set เป็นข้อมูลได้ 2 ลักษณะ ประกอบด้วย
 - i. ถ้าต้องการให้แสดงผลชุดข้อมูลที่เป็น outlier เลือก Marked samples (Outlier)
 - ii. ถ้าต้องการให้แสดงผลชุดข้อมูลที่ไม่มี outlier เลือก Unmarked samples (Training)
- 2) ใน create range จะเป็นการแสดงผลชุดข้อมูลใหม่จากการเลือก data matrix ที่ต้องการให้แสดงผลนั้นๆ ประกอบด้วย
 - iii. เลือก Create row set as new matrix from selected data matrices เป็นการสร้าง data matrix ใหม่ขึ้นมา
 - iv. ถ้าต้องการให้การแสดงผลชุดข้อมูลใหม่ปรากฏที่ขั้นตอน Preprocess ใดๆ ให้เลือกตาม column Select ตามตารางด้านล่าง

3) เมื่อกำหนดค่าแล้ว เลือก OK



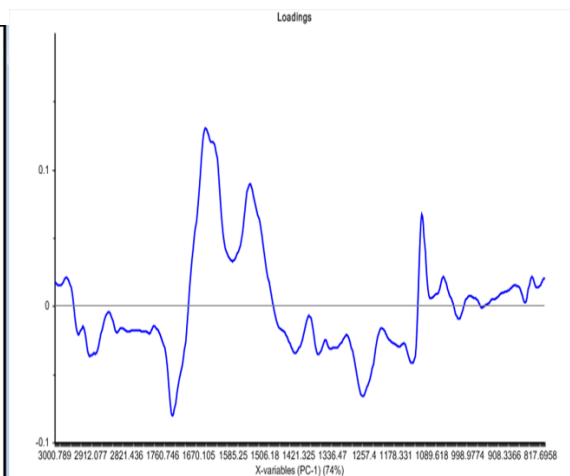
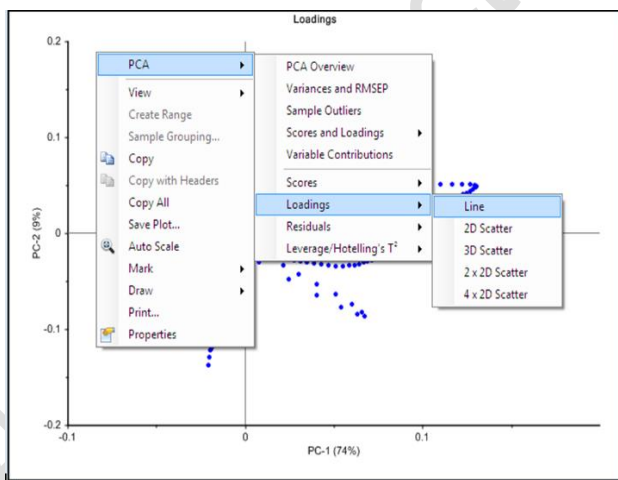
30. จะแสดงผลชุดข้อมูลตามที่ได้เลือกตั้งค่าไว้ จากตัวอย่าง เลือกให้ create data matrix ใหม่ ที่มี และ ไม่มี outlier ในชุดของ data set ที่เป็น original spectrum จะได้ผลการ create range แสดงดังภาพ
- โดย data matrix ที่ไม่มี outlier เรียกว่า Training set
 - และ data matrix ที่เป็น outlier เรียกว่า Outlier





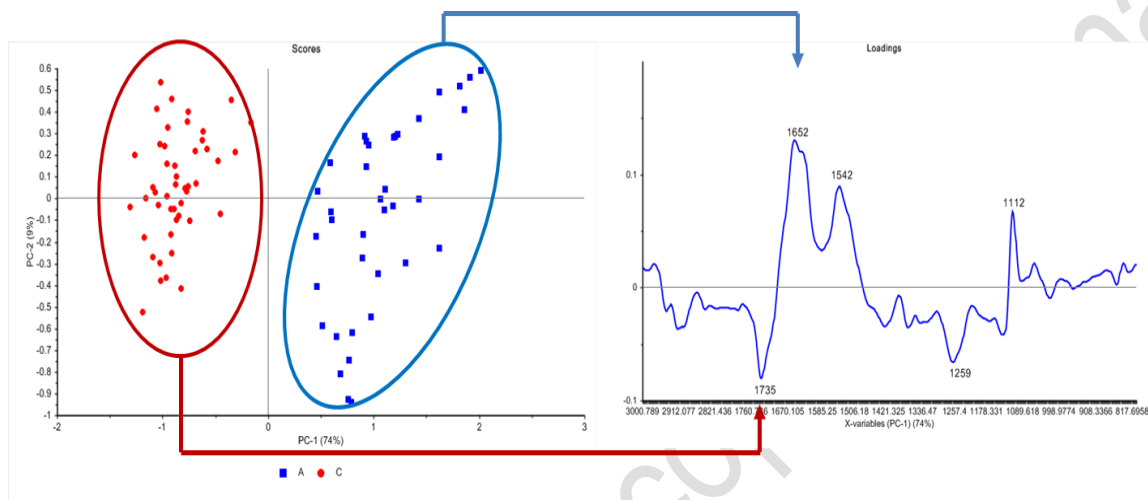
31. การอ่านผลการทำ PCA ประกอบกับ loading plot

- คลิกขวา เลือก PCA
- เลือก loadings
- เลือก line จะปรากฏหน้าต่างแสดงกราฟ loading ดังภาพ



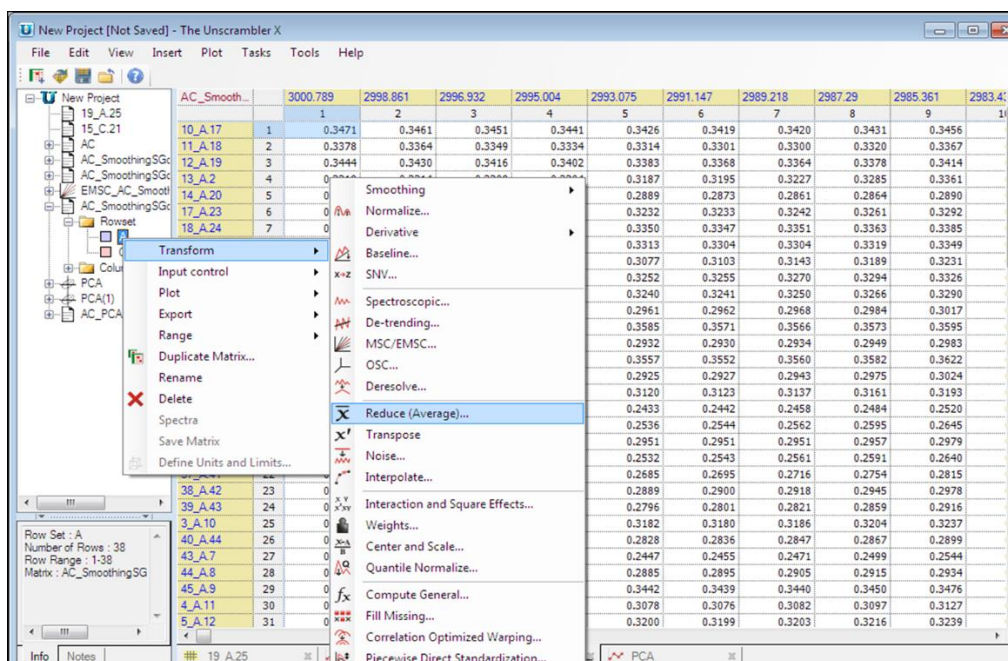
32. การอ่านผล PCA ทางแนวแกน PC1 ซึ่งแบ่งกลุ่มตัวอย่าง 2 กลุ่มออกเป็น

- ทางดานซ้ายมือของกลุ่มตัวอย่าง C คือตามแนวแกนค่าติดลบ (-) โดยจะสัมพันธ์กับ loading ตามแนวแกน Y ค่าลบ คือจะสัมพันธ์กับพีค 1735, 1259
- ทางด้านขวามือของกลุ่มตัวอย่าง A คือตามแนวแกนค่าบวก (+) โดยจะสัมพันธ์กับ loading ตามแนวแกน Y ค่าบวก คือจะสัมพันธ์กับพีค 1652, 1542, 1112



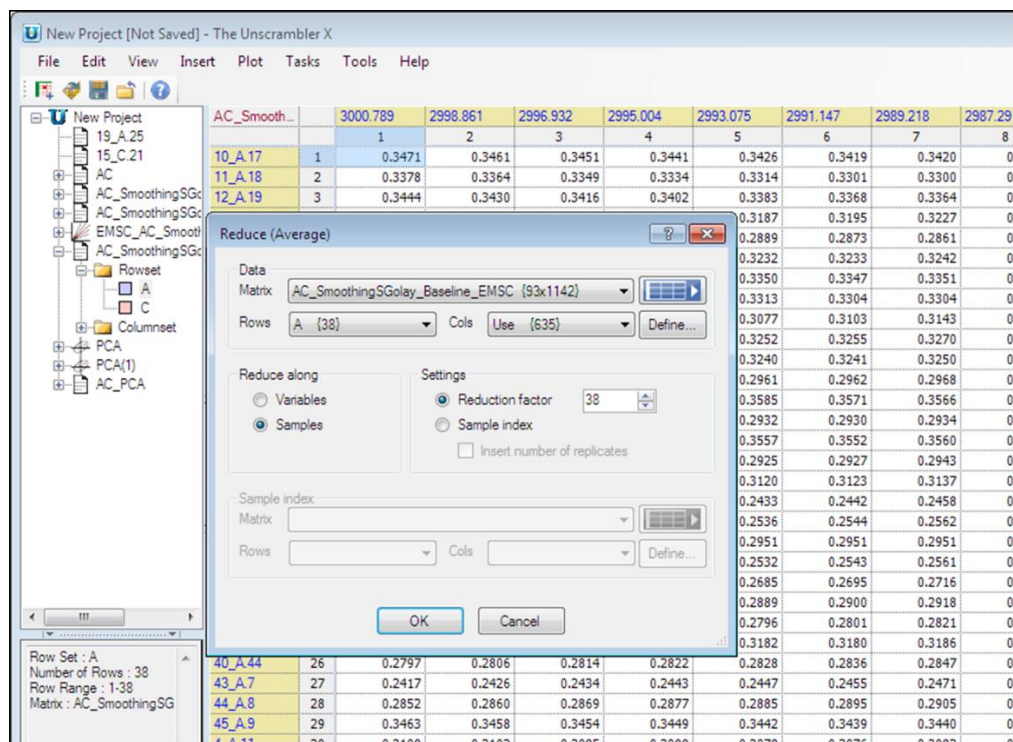
33. และเมื่อต้องการดูความแตกต่างของแต่ละกลุ่มตัวอย่าง สามารถทำได้โดยการเฉลี่ย spectrum ของแต่ละ group เพื่อเปรียบเทียบ

- คลิกเลือก data set ที่ต้องการเฉลี่ย spectrum ในตัวอย่าง เลือก average spectrum จาก data set ชุดที่ผ่านการทำ data preprocessing มาแล้ว
- คลิกขวา เลือก transform
- เลือก reduce average



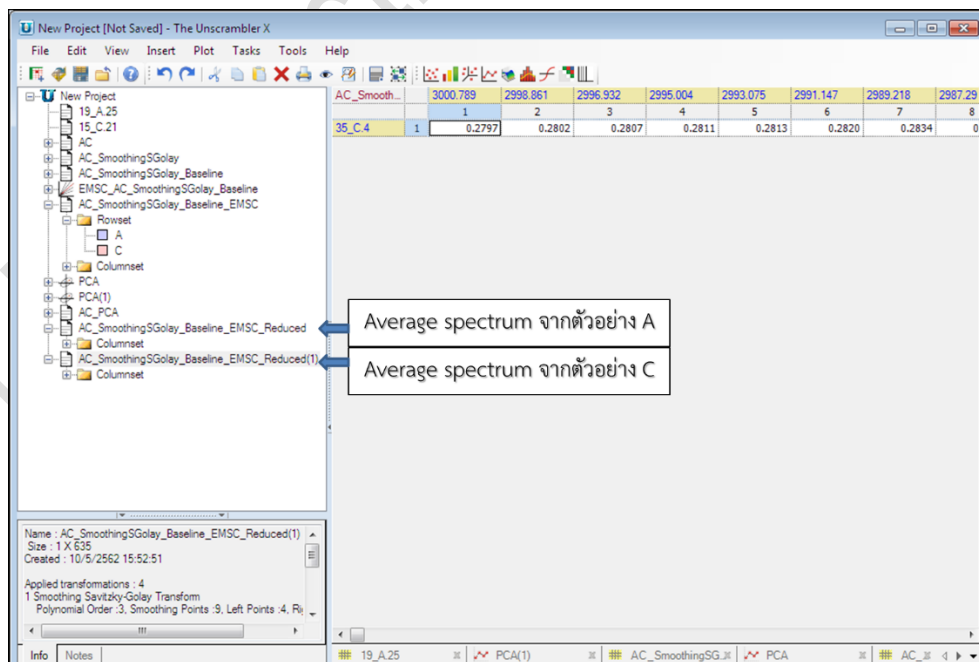
d. จะปรากฏหน้าต่างให้กำหนดค่า parameter โดยให้กำหนดค่าดังนี้

- i. Rows : เลือกกลุ่มตัวอย่างที่ต้องการเฉลี่ย spectrum
- ii. Cols : เลือก USE
- iii. Reduce along : เลือก samples
- iv. Setting : เลือก reduce factor แล้วใส่ตัวเลขเท่ากับจำนวน spectrum ของตัวอย่างนั้น ในตัวอย่าง กลุ่ม A มี spectrum ทั้งหมด 38 spectrum จึงใส่ค่า reduce factor เท่ากับ 38

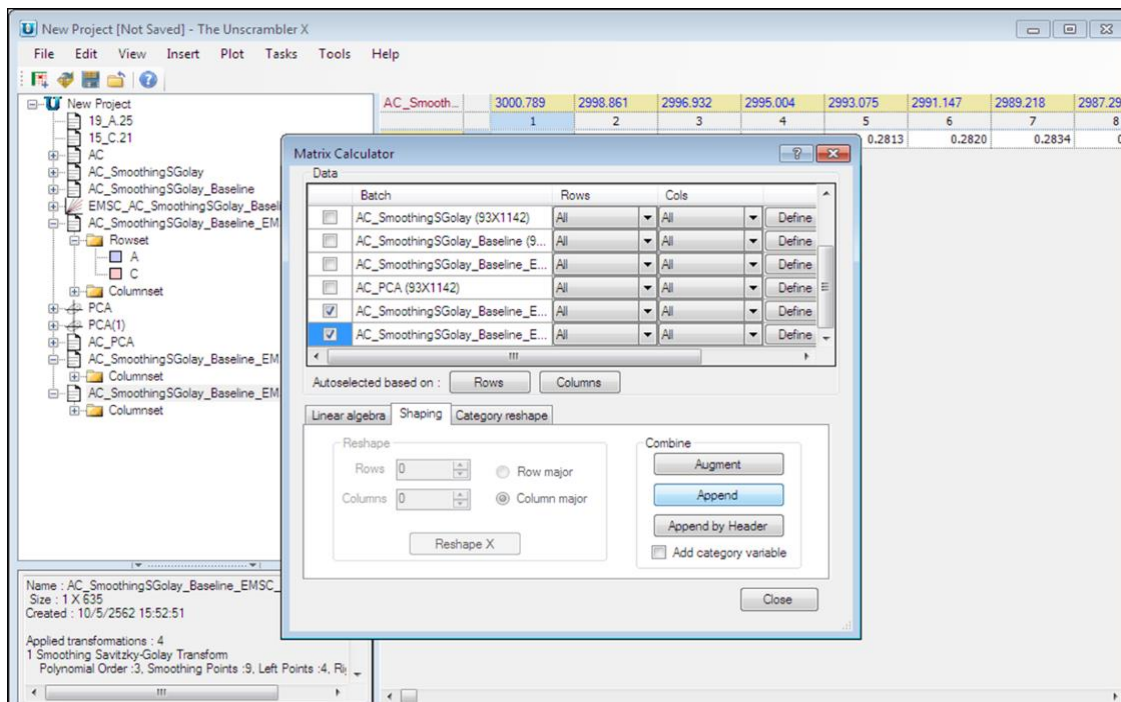


v. กด OK

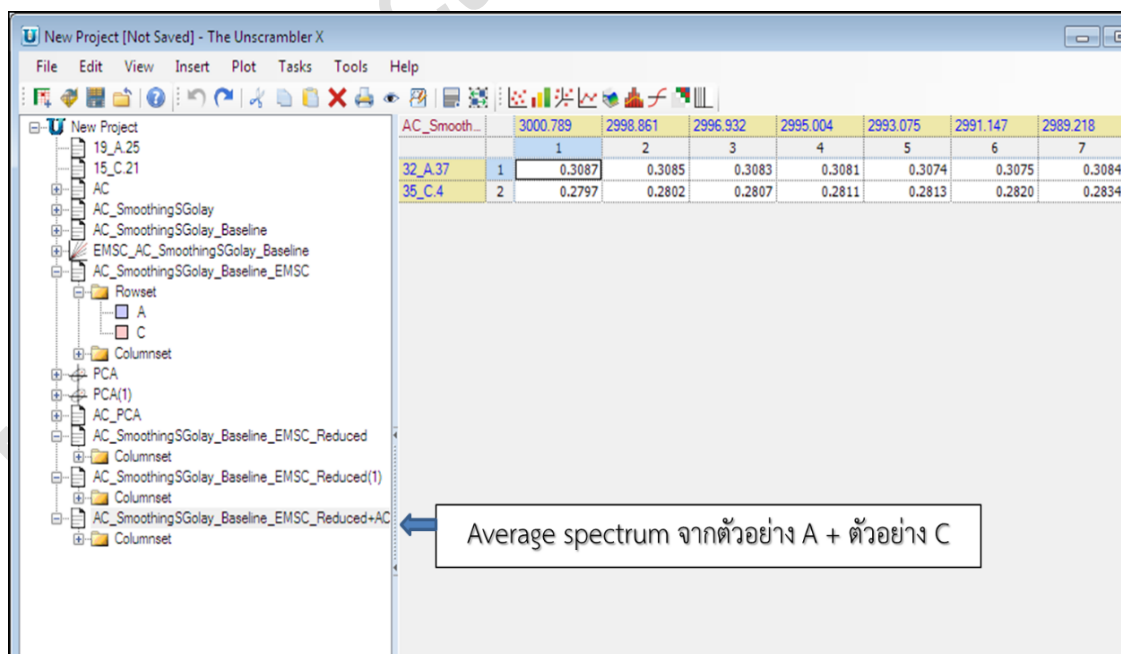
vi. และ average spectrum กับตัวอย่างกลุ่ม C โดยทำเหมือนวิธีข้อ i-v จะได้ spectrum ที่ผ่านการ average แล้วของตัวอย่าง A และตัวอย่าง C



34. เมื่อต้องการ plot กราฟเพื่อดูลักษณะ spectrum เปรียบเทียบ 2 ตัวอย่าง ให้รวมไฟล์ตัวอย่าง A และ C เข้าด้วยกันก่อน โดยการทำ matrix calculator (วิธีเหมือนข้อ 5)



35. จะได้ data set ใหม่ที่มีข้อมูล average spectrum ของตัวอย่าง A + ตัวอย่าง C



36. เมื่อต้องการ plot graph เพื่อดู spectrum เปรียบเทียบ ให้เลือกที่ data set ที่ต้องการ plot

- คลิกขวา เลือก Plot
- เลือก Line จะปรากฏหน้าต่าง Line Plot-Scope
- กำหนด matrix ที่ต้องการ Plot
- เลือก Sample
- เลือก OK จะปรากฏหน้าต่างแสดงกราฟ spectrum ที่ผ่านการ average แล้วเทียบตัวอย่าง A และตัวอย่าง C ซึ่งพิกที่ปรากฏมีความสัมพันธ์กันกับผลที่ได้จาก loading plot โดยจาก spectrum เกลี่ย จะเห็นได้ว่า ตัวอย่าง A จะแสดง intensity ของพิกบริเวณ 1652, 1542 และ 1112 มากกว่าตัวอย่าง C ในทางตรงข้าม ตัวอย่าง C จะแสดง intensity ของพิกบริเวณ 1736 และ 1249 มากกว่าตัวอย่าง A

