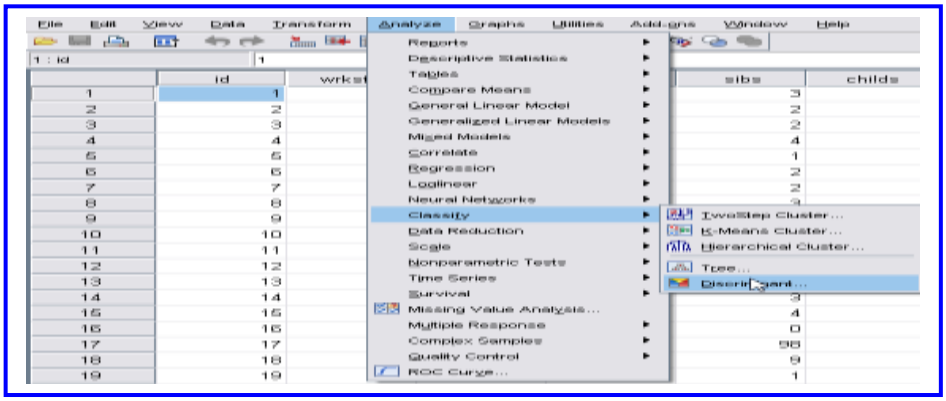
Discriminant Analysis Workshop #2-Stepwise method

ในเอกสารวิชาการนี้จะเป็นการแสดงตัวอย่างการใช้ Discriminant analysis เพื่อแยกแยะจัดเข้ากลุ่มบุคคลที่โหวตให้กับพรรคการเมืองหนึ่ง(vote92) โดยใช้ตัวแปรอิสระอันประกอบด้วยเพศ (sex) อายุ (age) ระดับการศึกษาสูงสุด (edu) รายได้ (rincome91) และความคิดเห็นเกี่ยวกับการเมือง (polviews) ตารางด้านล่างแสดงตัวอย่างข้อมูลที่จะใช้ DA ในการจัดกลุ่ม

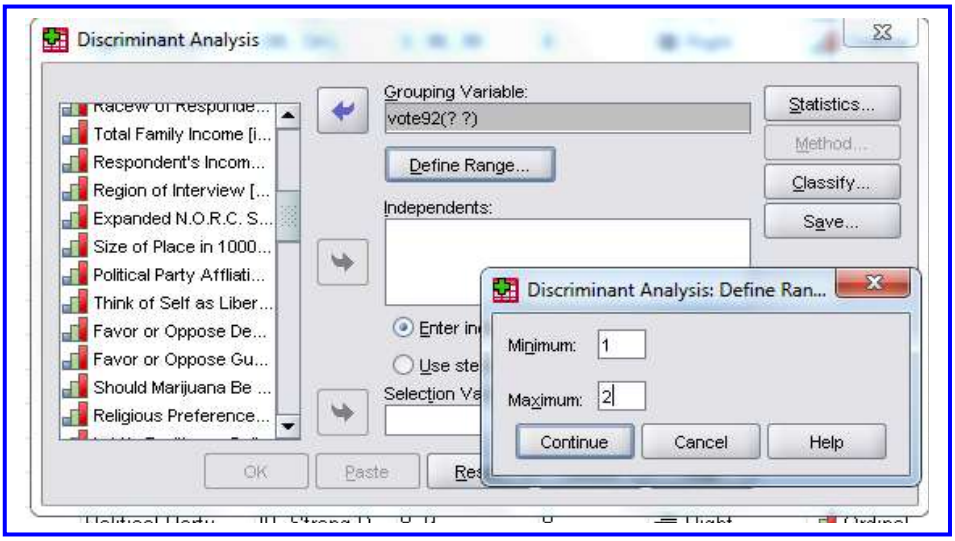
age	birthmo	zodiac	educ	degree	padeg	madeg	sex
43	5	2	11	1	7	0	1
44	8	6	16	3	0	0	1
43	2	11	16	3	1	0	2
45	99	99	15	1	1	8	2
78	10	7	17	4	1	2	2
83	3	12	11	1	0	0	1
55	10	7	12	1	0	0	2
75	11	9	12	1	0	0	1
31	7	4	18	4	1	1	1
54	3	12	18	4	0	1	2
29	4	2	18	4	0	0	2
23	10	8	15	1	7	2	2
61	99	99	12	1	1	1	2
63	3	1	4	0	7	0	2
33	3	12	10	0	8	8	2
36	11	8	14	1	7	1	2
39	3	12	8	0	9	0	2
55	1	10	15	1	1	0	1
55	9	7	16	3	1	0	2
34	4	2	16	3	0	1	1
36	6	3	14	2	3	3	2
44	8	5	18	3	1	1	2
80	5	2	18	4	7	0	1
32	7	5	16	3	3	3	1
37	9	6	5	0	0	0	1
49	9	6	16	3	3	3	2

rincom91	region	xnorcsiz	size	partyid	vote92	polviews
17	2	1	7072	5	1	5
18	2	1	7072	0	1	1
18	2	1	7072	5	1	6
22	2	1	7072	5	1	4
0	2	1	7072	3	1	3
0	2	1	7072	1	1	2
15	2	1	7072	0	1	2
0	2	1	7072	1	2	4
20	2	1	7072	6	1	5
2	2	1	7072	0	1	4
16	2	1	7072	1	2	1
3	2	1	7072	1	1	2
8	2	1	7072	1	3	2
0	2	1	7072	1	1	4
9	2	1	7072	3	2	6
10	2	1	7072	1	4	3
0	2	1	7072	5	2	4
22	2	1	7072	3	3	4
14	2	5	29	0	1	3
21	2	5	29	1	1	4
15	2	5	29	0	1	3
18	2	5	29	1	1	4
0	2	5	29	0	1	1
11	9	1	2967	0	3	3
4	9	1	2967	0	1	5
15	9	1	2967	6	1	5

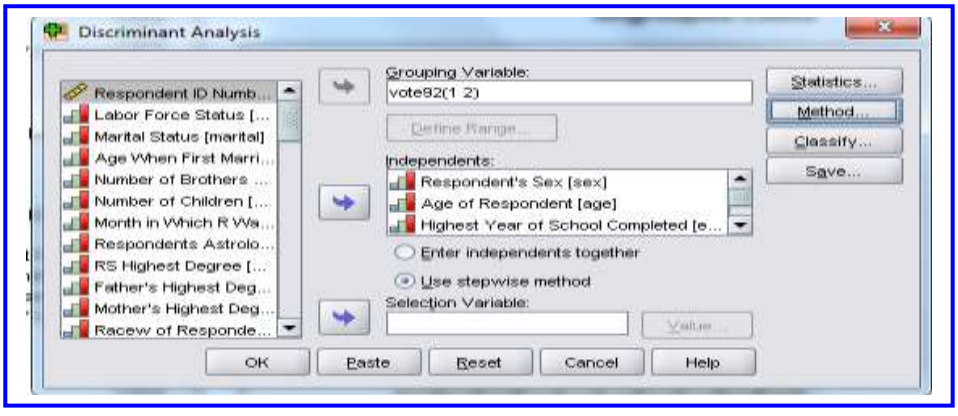
- คลิ๊ก Analyze>Classify>Discriminant



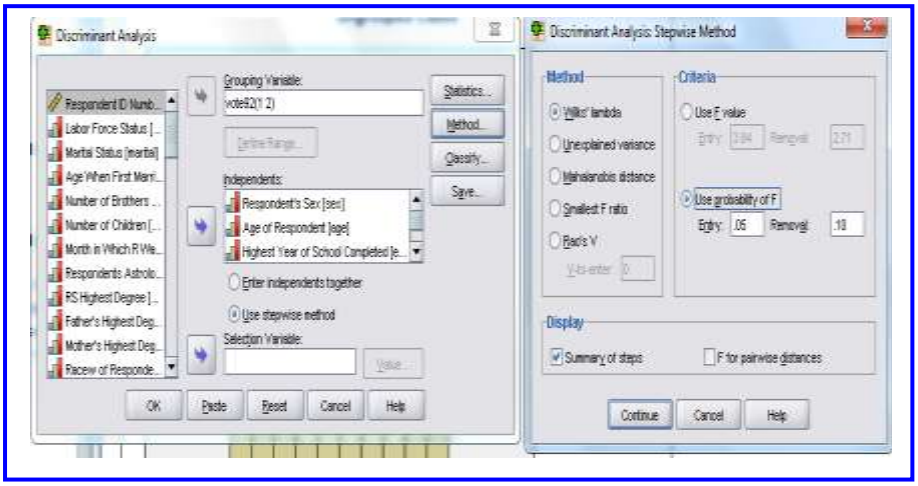
●คลิ๊กเลือกตัวแปรตาม(ในที่นี้ vote92) ให้เข้าไปอยู่ในช่อง Grouping Variable และคลิ๊กปุ่ม Define Range และใส่ค่าต่ำสุดเป็น1(โหวตให้) ส่วนค่าสูงสุดเป็น 2(ไม่ได้โหวตให้)



●คลิ๊กเลือกเพศ(sex) อายุ (age) การศึกษาสูงสุด (edu) รายได้ (rincome91) ความคิดเห็นทางการเมือง (polviews) ให้ไปอยู่ในกล่อง Independents คลิ๊กเลือกปุ่ม Use stepwise method



•คลิ๊กที่ปุ่ม Method ในส่วนของ Method เลือก Wilks’ Lambda ในส่วน Display คลิ๊ก Summary of steps ในส่วน Criteria เลือก Use probability of F (ส่วนตัวเลือก Entry และ Remove ไม่เปลี่ยนแปลง)



•คลิ๊กที่ปุ่ม Classify ในส่วน Prior Probabilities เลือก Compute from group sizes ในส่วน Display เลือก Casewise results และ Summary table ในส่วน Use Covariance Matrix เลือก Separate-groups ในส่วน Plots เลือก Separate-groups

ต่อไปนี้เป็น computer output พร้อมทั้งการตีความหมาย

Tests of Equality of Group Means

	Wilks' Lambda	F	df1	df2	Sig.
Respondent's Sex	1.000	.155	1	940	.694
Age of Respondent	.963	35.755	1	940	.000
Highest Year of School Completed	.919	82.767	1	940	.000
Respondent's Income	.954	45.170	1	940	.000
Think of Self as Liberal or Conservative	1.000	.268	1	940	.605

จาก Wilks’ Lambda ตัวแปรอิสระที่ช่วยในการแยกแยะจัดกลุ่มได้ดีคือ อายุ(age) การศึกษาสูงสุด (edu) และรายได้ (rincome91)

Box's Test of Equality of Covariance

Log Determinants

Voting in 1992 Election	Rank	Log Determinant
voted	3	10.253
did not vote	3	10.083
Pooled within-groups	3	10.219

The ranks and natural logarithms of determinants printed are those of the group covariance matrices.

Test Results

Box's M		9.392
F	Approx.	1.558
	df1	6
	df2	1.329E6
	Sig.	.155

Tests null hypothesis of equal population covariance matrices.

Box’s M test แสดงให้เห็นว่า Population covariance matrices เท่ากัน

Stepwise

Variables Entered/Removed[a,b,c,d]

Step	Entered	Wilks' Lambda				Exact F			
		Statistic	df1	df2	df3	Statistic	df1	df2	Sig.
1	Highest Year of School Completed	.919	1	1	940.000	82.767	1	940.000	.000
2	Age of Respondent	.870	2	1	940.000	70.237	2	939.000	.000
3	Respondent's Income	.865	3	1	940.000	48.941	3	938.000	.000

At each step, the variable that minimizes the overall Wilks' Lambda is entered.

a. Maximum number of steps is 10.

b. Maximum significance of F to enter is .05.

c. Minimum significance of F to remove is .10.

d. F level, tolerance, or VIN insufficient for further computation.

แสดงให้เห็นระดับการศึกษาสูงสุด (edu)เข้ามาในstep ที่ 1 อายุ(age) เข้ามาใน stepที่ 2 และรายได้ (rincome91) เข้ามาในstep ที่ 3

Variables in the Analysis

Step		Tolerance	Sig. of F to Remove	Wilks' Lambda
1	Highest Year of School Completed	1.000	.000	
2	Highest Year of School Completed	.972	.000	.963
	Age of Respondent	.972	.000	.919
3	Highest Year of School Completed	.861	.000	.931
	Age of Respondent	.925	.000	.904
	Respondent's Income	.864	.018	.870

แสดงข้อมูลเกี่ยวกับ Wilks' Lambda, Sig. of F ที่จะ entry ตลอดจน remove

Variables Not in the Analysis

Step		Tolerance	Min. Tolerance	Sig. of F to Enter	Wilks' Lambda
0	Respondent's Sex	1.000	1.000	.694	1.000
	Age of Respondent	1.000	1.000	.000	.963
	Highest Year of School Completed	1.000	1.000	.000	.919
	Respondent's Income	1.000	1.000	.000	.954
	Think of Self as Liberal or Conservative	1.000	1.000	.605	1.000
1	Respondent's Sex	1.000	1.000	.835	.919
	Age of Respondent	.972	.972	.000	.870
	Respondent's Income	.909	.909	.000	.904
	Think of Self as Liberal or Conservative	.994	.994	.252	.918
2	Respondent's Sex	1.000	.972	.646	.870
	Respondent's Income	.864	.861	.018	.865
	Think of Self as Liberal or Conservative	.986	.964	.636	.870
3	Respondent's Sex	.926	.861	.253	.863
	Think of Self as Liberal or Conservative	.986	.858	.664	.864

Wilks' Lambda

Step	Number of Variables	Lambda	df1	df2	df3	Exact F			
						Statistic	df1	df2	Sig.
1	1	.919	1	1	940	82.767	1	940.000	.000
2	2	.870	2	1	940	70.237	2	939.000	.000
3	3	.865	3	1	940	48.941	3	938.000	.000

แสดงให้เห็นว่าในแต่ละ step จำนวนตัวแปรอิสระที่นำเข้ามาล้วนมีส่วนสำคัญในการช่วยแยกแยะจัดกลุ่ม

Summary of Canonical Discriminant

Eigenvalues

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	.157[a]	100.0	100.0	.368

a. First 1 canonical discriminant functions were used in the analysis.

Wilks' Lambda

Test of...	Wilks' Lambda	Chi-square	df	Sig.
1	.865	136.478	3	.000

Canonical correlation ยกกำลังสองแสดงให้เห็นว่าร้อยละ 13.5 ของความผันผวนในตัวแปรตามสามารถแยกแยะได้ด้วยกลุ่มตัวแปรอิสระ

Wilks' Lambda แสดงให้เห็นว่าmodelที่ใช้มีนัยสำคัญ และสามารถใช้ในการแยกแยะจัดกลุ่มได้

Standardized Canonical Discriminant Function Coefficients

	Function 1
Age of Respondent	.588
Highest Year of School Completed	.780
Respondent's Income	.226

Standardized Canonical Discriminant Function Coefficients แสดงให้เห็นระดับความสำคัญของตัวแปรอิสระในการแยกแยะจัดกลุ่ม ในที่นี้ระดับการศึกษาสูงสุด(edu) เป็นตัวแปรอิสระที่ช่วยในการแยกแยะจัดกลุ่มได้สูงที่สุดในบรรดาตัวแปรอิสระทั้งหลายที่นำเข้ามา

Structure Matrix

	Function 1
Highest Year of School Completed	.750
Respondent's Income	.554
Age of Respondent	.493
Respondent's Sex[a]	-.065
Think of Self as Liberal or Conservative[a]	.004

Pooled within-groups correlations between discriminating variables and standardized canonical discriminant functions
Variables ordered by absolute size of correlation within function.

a. This variable not used in the analysis.

Structure matrix แสดงให้เห็นความสัมพันธ์ระหว่างตัวแปรอิสระกับ discriminant function ในที่นี้ระดับการศึกษาสูงสุด(edu) มีความสัมพันธ์สูงสุดกับ discriminant function

Canonical Discriminant Function Coefficients

	Function 1
Age of Respondent	.048
Highest Year of School Completed	.292
Respondent's Income	.041
(Constant)	-6.476

Unstandardized coefficients

Canonical Discriminant Function Coefficients ให้ข้อมูลในการสร้าง Discriminant function ที่จะช่วยในการแยกแยะจัดกลุ่มในอนาคต ในที่นี้รูปแบบของ discriminant function คือ

L=-6.476+.048(age)+.292(edu)+.041(rincome91)

Functions at Group Centroids

Voting in 1992 ...	Function 1
voted	.235
did not vote	-.665

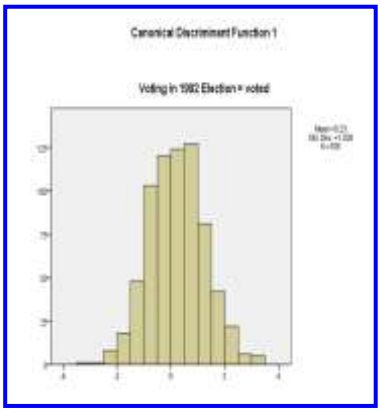
Unstandardized canonical discriminant functions evaluated at group means

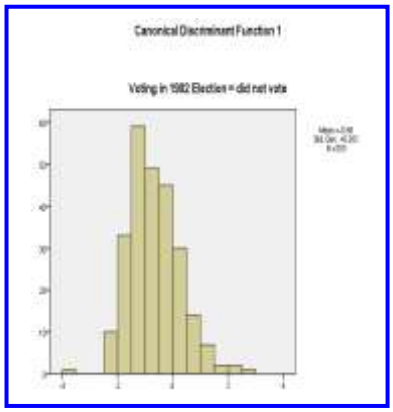
Functions at Group Centroids แสดงให้เห็น mean discriminant scores ของแต่ละกลุ่มในตัวแปรตาม ในที่นี้พวกที่โหวตให้ มี mean discriminant score= 0.235 ในขณะที่ของผู้ไม่ได้โหวตให้=-.665 อนึ่ง เนื่องจากค่าของ mean discriminant scores ของสองกลุ่มมีความแตกต่างกันค่อนข้างมาก การแยกแยะจัดกลุ่มจะมีความผิดพลาดน้อย

Casewise Statistics

	Case Number	Actual Group	Highest Group: Predicted Group	Highest Group: P(D=d | G=g): p	Highest Group: P(D=d | G=g): df	Highest Group: P(G=g | D=d)	Highest Group: Squared Mahalanobis Distance to Centroid	Second Highest Group: Group	Second Highest Group: P(G=g | D=d)	Second Highest Group: Squared Mahalanobis Distance to Centroid	Discriminant Scores: Function 1
Original	1	1	1	.464	1	.661	.535	2	.339	.826	-.517
	2	1	1	.438	1	.912	.599	2	.088	3.435	1.031
	3	1	1	.467	1	.907	.530	2	.093	3.245	.983
	4	1	1	.498	1	.904	.486	2	.096	3.123	.952
	7	1	1	.979	1	.808	.061	2	.192	1.827	.262
	9	1	1	.412	1	.917	.674	2	.083	3.634	1.079
	10	1	1	.246	1	.946	1.347	2	.054	5.235	1.439
	11	2	1**	.570	1	.889	.322	2	.111	2.628	.819
	12	1	1	.278	1	.598	1.176	2	.410	.855	-.880
	13	ungrouped	1	.992	1	.807	.061	2	.193	1.818	.259
	15	2	2	.298	1	.536	1.080	1	.461	3.238	-1.616
	16	ungrouped	1	.627	1	.711	.236	2	.289	.192	-.284
	18	ungrouped	1	.246	1	.946	1.345	2	.054	5.231	1.438
	19	1	1	.262	1	.943	1.259	2	.057	5.037	1.389
	20	1	1	.865	1	.871	.187	2	.129	2.158	.680
	21	1	1	.778	1	.751	.081	2	.249	.441	-.057
	22	1	1	.168	1	.958	1.881	2	.042	6.208	1.615
	24	ungrouped	1	.958	1	.792	.084	2	.208	.834	.171
	25	1	2**	.808	1	.698	7.037	1	.312	10.467	-3.082
	26	1	1	.378	1	.923	.783	2	.077	3.911	1.145
	27	1	1	.368	1	.925	.812	2	.075	3.994	1.181
	28	1	1	.542	1	.894	.372	2	.106	2.795	.862
	29	1	1	.472	1	.864	.518	2	.336	.838	-.585

Casewise Statistics(แสดงบางส่วน) แสดงรายละเอียดการแยกแยะจัดกลุ่มของตัวแปรตามโดยใช้ข้อมูลที่ได้จาก Discriminant function





แสดง histogram ให้เห็นผลจากการใช้ discriminant functionในการแยกแยะจัดกลุ่ม

ใน model ที่ DA มีประสิทธิภาพในการแยกแยะจัดกลุ่มสูง กราฟรูปแท่งจะอยู่สูงและกระจุกอยู่ใกล้ค่าเฉลี่ย ส่วนปลายจะมีกราฟรูปแท่งเล็ก ๆ

Classification Results[a]

		Voting in 1992 Election	Predicted Group Membership: voted	Predicted Group Membership: did not vote	Total
Original	Count	voted	670	36	706
		did not vote	198	55	253
		Ungrouped cases	25	4	29
	%	voted	94.9	5.1	100.0
		did not vote	78.3	21.7	100.0
		Ungrouped cases	86.2	13.8	100.0

a. 75.6% of original grouped cases correctly classified.

ความถูกต้องในการแยกแยะจัดกลุ่ม
= 75.6%
[(670+55)X100/(706+253)]

Contribution this issue: ดร. ดนัย ปัตตพงศ์

อยากเรียนรู้การนำสถิติข้างต้นนี้ไปใช้ในการวิจัยระดับสารนิพนธ์ (independent study) วิทยานิพนธ์ (thesis) ดุษฎีนิพนธ์(dissertation) ปรึกษาได้ที่ dpattaphongse@gmail.com

* ผู้แต่ง MBA's Made Easy (160+ issues) เอกสารวิชาการด้านศาสตร์การบริหารธุรกิจที่ช่วยให้ธุรกิจสามารถยืนหยัดและอยู่รอดได้ในภาวะที่โลกเปลี่ยนแปลงอยู่ตลอดเวลา

* ผู้พัฒนา FINALYSIS… a dedicated software สำหรับให้บริการนักธุรกิจที่ต้องการวิเคราะห์ความเป็นไปได้ทางการเงินของโครงการพัฒนาอสังหาริมทรัพย์ (บ้านจัดสรร/จัดสรรที่ดินเพื่อการอุตสาหกรรม/อาคารชุด/อาคารสำนักงานให้เช่า) โรงแรม โรงพยาบาลเอกชน ห้างสรรพสินค้า โรงงานน้ำตาล โรงงานกระดาษ โรงไฟฟ้าชีวมวล ฯลฯ ได้เห็นตัวเลขก่อนโครงการเกิด หลีกเลี่ยงความผิดพลาดเป็นร้อยเป็นพันล้านหากเกิดการลงทุนจริง(กำหนด DEBUT 1 เมษายน 2569)

* ผู้แต่งหนังสือ"การวิเคราะห์ความเป็นไปได้ทางการเงินและการจัดวงเงินเครดิตของโครงการลงทุน"ประกอบด้วยตัวอย่างของธุรกิจจริงที่ไม่เปิดเผยชื่อนับ 100 บริษัท ครอบคลุมอุตสาหกรรม 24 อุตสาหกรรม

*Co-developer ซอฟแวร์ en◉gex◉cel® สำหรับใช้ทดสอบ/เรียนรู้ศัพท์(ประกอบด้วยแบบฝึกหัดและเฉลยกว่า 90 บทครอบคลุมศัพท์ระดับ SAT/IELTS/TOEFL กว่า 12,000 คำ) และไวยากรณ์อังกฤษ (ประกอบด้วยแบบฝึกหัดและเฉลยกว่า 160 บทหรือกว่า 10,000 ข้อครอบคลุมเนื้อหาระดับอุดมศึกษาและTOEFL) มาพร้อมกับไฟล์เสียง/ไฟล์ข้อมูล/ฯลฯ อีกมาก(กำหนด DEBUT 1 เมษายน 2569)