

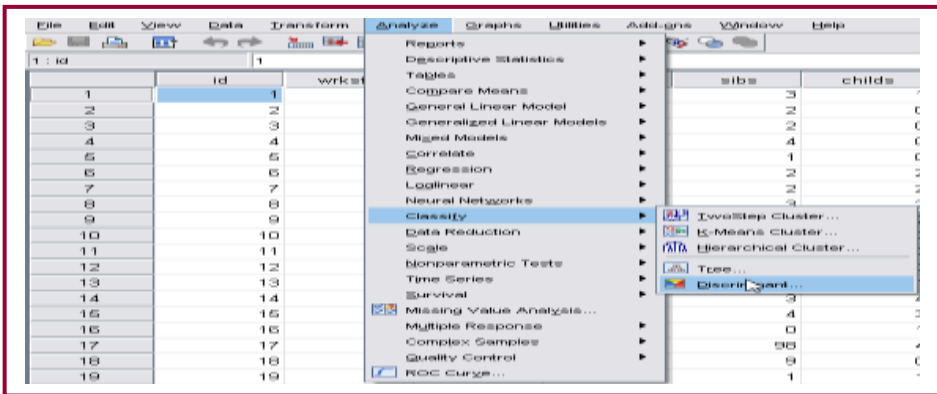


Discriminant Analysis Workshop #1-Enter method

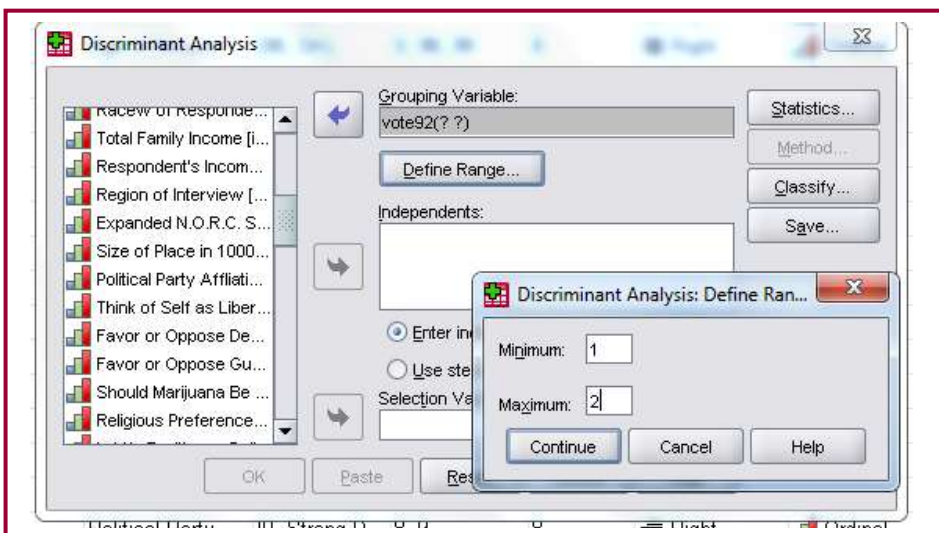
ในเอกสารวิชาการนี้จะเป็นการแสดงตัวอย่างการใช้ Discriminant analysis เพื่อแยกแยะจัดเข้ากลุ่มบุคคลที่โหวตให้กับพรรคการเมืองหนึ่ง (vote92) โดยใช้ตัวแปรอิสระอันประกอบด้วยเพศ (sex) อายุ (age) ระดับการศึกษาสูงสุด (edu) รายได้ (rincome91) และความคิดเห็นเกี่ยวกับการเมือง (polviews) ตารางด้านล่างแสดงตัวอย่างข้อมูลที่จะใช้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

age	birthmo	zodiac	educ	degree	padeg	madeg	sex	rincome91	region	xmomsiz	size	partyid	vote92	polviews
43	5	2	11	1	7	0	1	17	2	1	7072	5	1	5
44	8	6	16	3	0	0	1	18	2	1	7072	0	1	1
43	2	11	16	3	1	0	2	18	2	1	7072	5	1	6
45	99	99	15	1	1	8	2	22	2	1	7072	5	1	4
78	10	7	17	4	1	2	2	0	2	1	7072	3	1	3
83	3	12	11	1	0	0	1	0	2	1	7072	1	1	2
55	10	7	12	1	0	0	2	15	2	1	7072	0	1	2
75	11	9	12	1	0	0	1	0	2	1	7072	1	2	4
31	7	4	18	4	1	1	1	20	2	1	7072	6	1	5
54	3	12	18	4	0	1	2	2	2	1	7072	0	1	4
29	4	2	18	4	0	0	2	16	2	1	7072	1	2	1
23	10	8	15	1	7	2	2	3	2	1	7072	1	1	2
61	99	99	12	1	1	1	2	8	2	1	7072	1	3	2
63	3	1	4	0	7	0	2	0	2	1	7072	1	1	4
33	3	12	10	0	8	8	2	9	2	1	7072	3	2	6
36	11	8	14	1	7	1	2	10	2	1	7072	1	4	3
38	3	12	8	0	9	0	2	0	2	1	7072	5	2	4
55	1	10	15	1	1	0	1	22	2	1	7072	3	3	4
55	9	7	16	3	1	0	2	14	2	5	29	0	1	3
34	4	2	16	3	0	1	1	21	2	5	29	1	1	4
36	6	3	14	2	3	3	2	15	2	5	29	0	1	3
44	8	5	18	3	1	1	2	18	2	5	29	1	1	4
80	5	2	18	4	7	0	1	0	2	5	29	0	1	1
32	7	5	16	3	3	3	1	11	9	1	2967	0	3	3

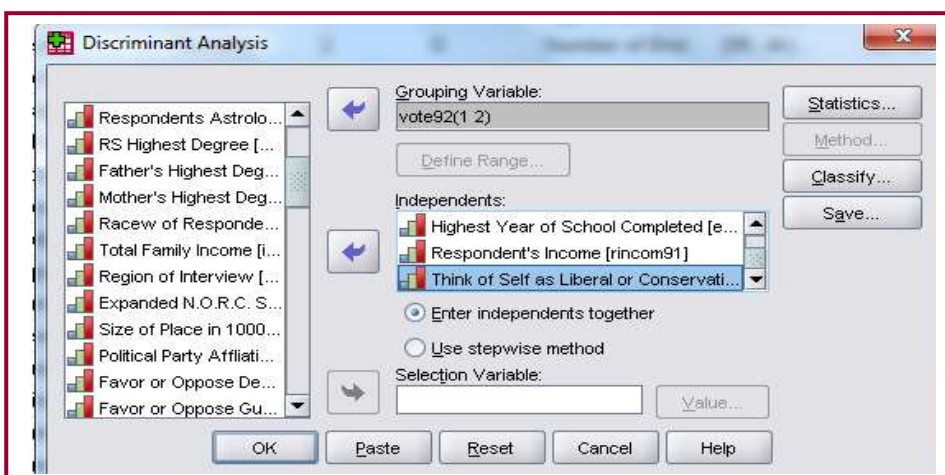
- คลิก Analyze>Classify>Discriminant



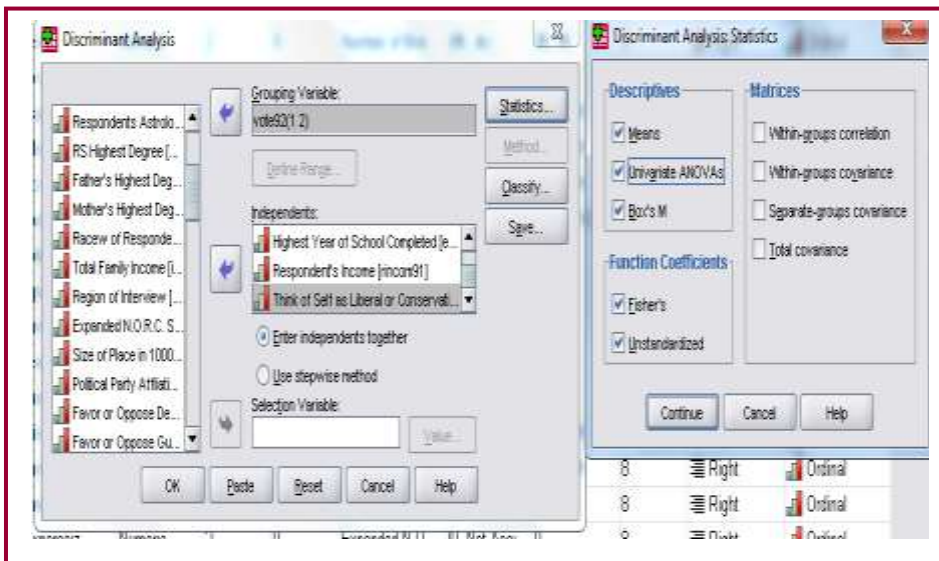
- คลิกเลือกตัวแปรตาม(ในที่นี้ vote92) ให้เข้าไปอยู่ในช่อง Grouping Variable และคลิกปุ่ม Define Range และใส่ค่าต่ำสุดเป็น 1(โหวตให้) ส่วนค่าสูงสุดเป็น 2(ไม่ได้โหวตให้)



- คลิกเลือกตัวแปรอิสระ(sex, age, educ, rincome91, polviews)ที่จะใช้ในการจัดกลุ่มเข้าไปอยู่ในช่อง Independents



- คลิกที่ปุ่ม Statistics ในส่วนของ Descriptives เลือก Means , Univariate Anovas, Box's M ในส่วน Function Coefficients เลือก Fisher's , Unstandardized คลิก Continue



- คลิกที่ปุ่ม 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) และ รายได้(rincome91)

Box's Test of Equality of Covariance

Log Determinants

Voting in 1992 Election	Rank	Log Determinant
voted	5	9.438
did not vote	5	8.815
Pooled within-groups	5	9.311

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

Test Results

Box's M		33.185
F	Approx.	2.195
	df1	15
	df2	8.784E5
	Sig.	.005

Tests null hypothesis of equal population covariance matrices.

Box's M test แสดงให้เห็นว่า Population covariance matrices ไม่เท่ากัน อย่างไรก็ตามในขนาดของกลุ่มตัวอย่างที่ใหญ่(ในกรณีนี้ n = 1500)ความแตกต่างเพียงเล็กน้อยมีผลทำให้ปฏิเสธnull hypothesis ได้ และเราจะมองข้ามผลนี้ไป

Eigenvalues

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	.158 ^a	100.0	100.0	.370

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

Wilks' Lambda

Test of ...	Wilks' Lambda	Chi-square	df	Sig.
1	.863	137.910	5	.000

Canonical correlation ยกกำลังสอง แสดงให้เห็นว่าร้อยละ 13.7 ของความผันผวนในตัวแปรตามสามารถแยกแยะได้ด้วยกลุ่มตัวแปรอิสระ Wilks' Lambda แสดงให้เห็นว่าmodel ที่ใช้นั้นยสำคัญ และสามารถใช้ในการแยกแยะจัดกลุ่มได้

Standardized Canonical Discriminant Function Coefficients

	Function	
	1	
Respondent's Sex		.108
Age of Respondent		.573
Highest Year of School Completed		.769
Respondent's Income		.256
Think of Self as Liberal or Conservative		.046

Standardized Canonical Discriminant Function Coefficients แสดงให้เห็นระดับความสำคัญของตัวแปรอิสระในการแยกแยะจัดกลุ่ม ในที่นี้ระดับการศึกษาสูงสุด(edu) เป็นตัวแปรอิสระที่ช่วยในการแยกแยะจัดกลุ่มได้สูงที่สุดในบรรดาตัวแปรอิสระทั้งหลาย

Structure Matrix

	Function	
	1	
Highest Year of School Completed		.745
Respondent's Income		.551
Age of Respondent		.490
Think of Self as Liberal or Conservative		.042
Respondent's Sex		.032

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

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

Canonical Discriminant Function Coefficients

	Function
	1
Respondent's Sex	.217
Age of Respondent	.046
Highest Year of School Completed	.288
Respondent's Income	.047
Think of Self as Liberal or Conservative	.035
(Constant)	-6.910

Unstandardized coefficients

Canonical Discriminant Function Coefficients ให้ข้อมูลในการสร้าง Discriminant function ที่จะช่วยในการแยกแยะจัดกลุ่มในอนาคต ในที่นี้ รูปแบบของ discriminant function คือ

$$L = -6.910 + 0.217(\text{sex}) + .046(\text{age}) + .288(\text{edu}) + .047(\text{rincome91}) + .035(\text{polview})$$

Functions at Group Centroids

Voting in 1992 ...	Function
	1
voted	.236
did not vote	-.669

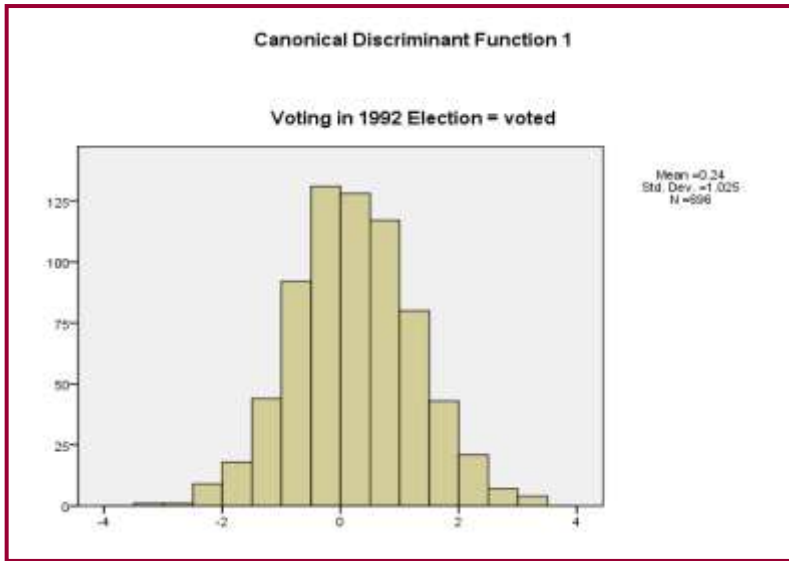
Unstandardized canonical discriminant functions evaluated at group means

Functions at Group Centroids แสดงให้เห็น mean discriminant scores ของแต่ละกลุ่มในตัวแปรตาม ในที่นี้พวกที่โหวตให้ มี mean discriminant score=0.236 ในขณะที่ ของผู้ไม่ได้โหวต ให้=-.669 อื่นๆเนื่องจากค่าของ mean discriminant scores ของสองกลุ่มมีความแตกต่างกันค่อนข้างมาก การแยกแยะจัดกลุ่มจะมีความผิดพลาดน้อย

Casewise Statistics

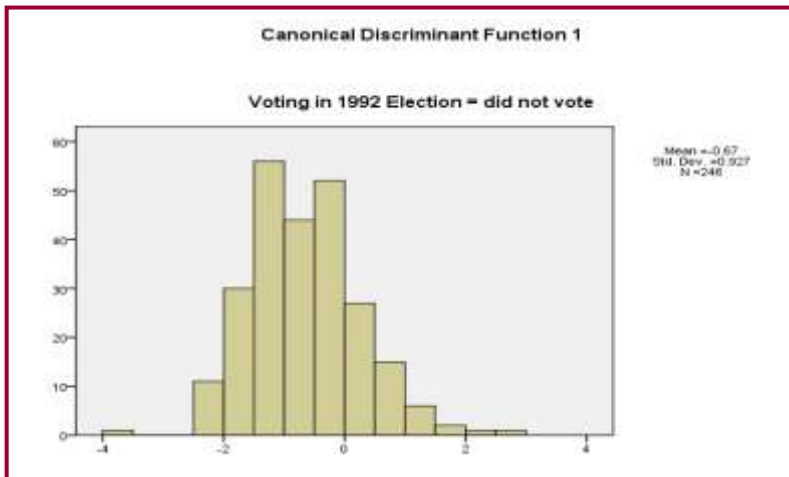
Original	Case Number	Actual Group	Predicted Group	Highest Group				Second Highest Group				Discriminant Scores
				P(D=1 D=1)		P(D=1 D=2)	Squared Mahalanobis Distance to Centroid	Group	P(D=2 D=2)		Squared Mahalanobis Distance to Centroid	
				n	#				n	#		
1	1	1	1	434	1	.006	.813	2	.245	.812	-.568	
2	1	1	1	565	1	.888	.332	2	.112	2.903	.826	
3	1	1	1	362	1	.924	.831	2	.076	3.939	1.171	
4	1	1	1	483	1	.917	.899	2	.083	3.814	1.060	
7	1	1	1	553	1	.815	.893	2	.185	1.996	.297	
9	1	1	1	438	1	.911	.892	2	.089	3.386	1.031	
10	1	1	1	341	1	.945	1.177	2	.055	5.172	1.438	
11	2	1	1	563	1	.888	.335	2	.111	2.912	.828	
12	1	1	1	274	1	.591	1.199	2	.409	.854	-.065	
13	angrased	1	1	391	1	.807	.990	2	.193	.879	.248	
15	2	2	2	485	1	.513	.893	1	.487	2.878	-1.440	
16	angrased	1	1	685	1	.725	.988	2	.275	.247	-.208	
18	angrased	1	1	281	1	.928	1.181	2	.072	4.700	1.348	
19	1	1	1	342	1	.944	1.170	2	.056	5.158	1.436	
20	1	1	1	718	1	.881	.731	2	.119	1.996	.827	
21	1	1	1	817	1	.768	.842	2	.232	.562	.026	
22	1	1	1	147	1	.981	2.184	2	.019	6.880	1.722	
24	angrased	1	1	837	1	.786	.848	2	.214	.540	.012	
25	1	2	2	107	1	.721	7.331	1	.279	11.106	-3.178	
26	1	1	1	311	1	.932	1.825	2	.068	4.394	1.274	
27	1	1	1	331	1	.931	.883	2	.069	4.298	1.252	
28	1	1	1	542	1	.883	.372	2	.117	2.726	.861	
29	1	1	1	478	1	.667	.522	2	.333	.832	-.504	
30	1	1	1	381	1	.801	.891	2	.199	.894	.212	
31	1	1	1	681	1	.732	.559	2	.268	.299	-.171	

Casewise Statistics (แสดงบางส่วน) แสดงรายละเอียดการแยกแยะจัดกลุ่มของตัวแปรตามโดยใช้ข้อมูลที่ได้จาก Discriminant function



แสดง histogram ให้เห็นผลจากการใช้ discriminant function (ในที่นี้มีเพียง 1 function) ในการแยกแยะจัดกลุ่ม

ใน model ที่ DA มีประสิทธิภาพในการแยกแยะจัดกลุ่มสูง กราฟรูปแท่งจะอยู่สูงและกระจุกอยู่ใกล้ค่าเฉลี่ย ส่วนปลายจะมีกราฟรูปแท่งเล็ก ๆ



ความถูกต้องในการแยกแยะจัดกลุ่ม = 76.1%

$$[(661+56) \times 100 / (696+246)]$$

Classification Results^a

		Voting in 1992 Election	Predicted Group Membership		Total
			voted	did not vote	
Original	Count	voted	661	35	696
		did not vote	190	56	246
		Ungrouped cases	21	2	23
%		voted	95.0	5.0	100.0
		did not vote	77.2	22.8	100.0
		Ungrouped cases	91.3	8.7	100.0

a. 76.1% of original grouped cases correctly classified.

Contribution this issue: ดร. ดนัย ปัตตพงษ์

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 ซอฟต์แวร์ 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)