

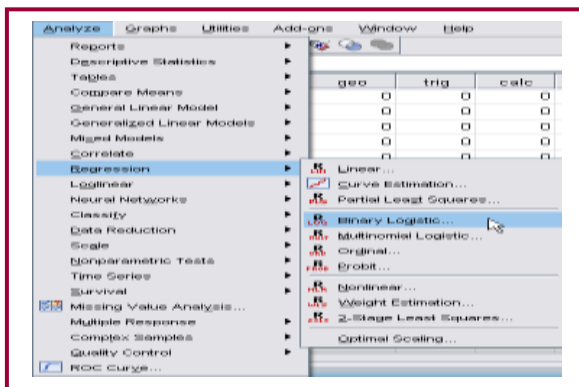
## Logistic Regression Workshop #3-Hierarchical method

ภาพด้านล่างแสดงให้เห็นลักษณะของข้อมูลที่เราต้องการพยากรณ์การเลือกเรียนวิชาพีชคณิต2 โดยอาศัยข้อมูลจากเพศของนักเรียน ระดับการศึกษาของบิดา ผลคะแนนที่ได้จากการทดสอบวิชาว่าด้วยลวดลาย ( mosaic test) และผลคะแนนที่ได้จากการทดสอบวิชาว่าด้วยการมองเห็น (visualization test)

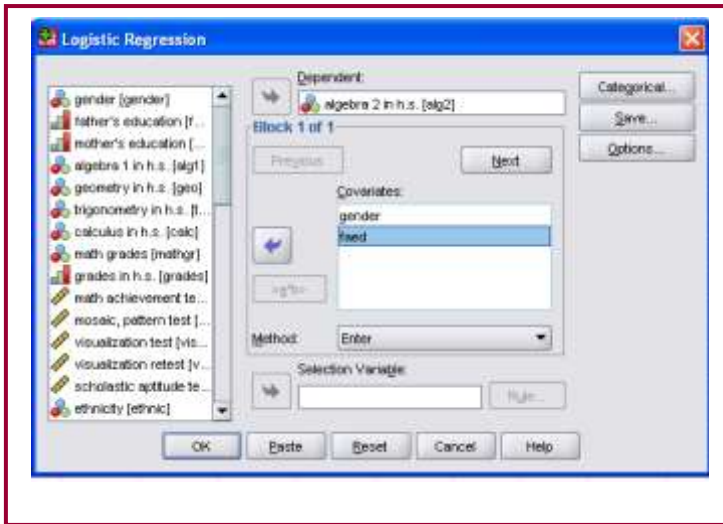
	gender	faed	maed	alg1	alg2	geo	trig	calc	mathgr	grades	mathach	mosaic	visual	visual2	satm
1	1	10	10	0	0	0	0	0	0	4	9.00	31.0	8.75	7.00	500
2	1	2	2	0	0	0	0	0	0	5	10.33	66.0	4.75	4.00	460
3	1	2	2	0	0	0	0	0	1	6	7.67	25.0	4.75	3.00	420
4	0	3	3	1	0	0	0	0	0	3	5.00	22.0	1.00	1.00	400
5	1	3	3	0	0	0	0	0	0	3	-1.67	17.5	2.25	2.00	450
6	1	3	2	0	0	0	0	0	1	5	1.00	23.5	1.00	0.00	260
7	0	9	6	1	1	1	1	0	0	6	12.00	28.5	2.50	2.00	480
8	1	5	3	1	0	0	0	0	0	4	8.00	29.5	3.50	3.00	430
9	1	3	3	1	0	0	0	0	1	7	13.00	28.0	3.50	2.00	490
10	0	8	2	0	0	0	0	0	0	5	3.67	27.5	3.75	3.00	300
11	0	3	4	1	1	1	1	1	1	6	21.00	27.0	11.00	9.60	600
12	0	8	9	1	1	1	1	1	1	8	23.67	26.5	4.75	4.00	700
13	1	2	2	1	0	0	0	0	0	5	4.00	13.0	1.00	1.00	380
14	0	6	3	1	0	0	0	0	0	2	9.00	18.0	1.00	0.00	450
15	1	2	3	1	1	0	0	0	0	3	5.33	26.0	4.75	4.00	400
16	0	3	3	1	1	1	1	0	1	7	19.67	33.0	13.50	9.50	690
17	0	3	3	1	0	1	0	0	0	7	7.67	39.0	-0.25	1.00	430
18	0	9	7	1	0	1	0	0	0	5	14.33	27.0	6.00	6.00	500
19	0	3	3	1	1	1	0	0	1	6	14.33	26.0	9.75	9.50	510
20	1	8	3	1	1	1	1	0	1	8	19.67	26.0	6.75	9.50	500
21	0	5	5	1	0	1	0	0	0	5	17.00	41.0	-0.25	1.00	580
22	1	2	3	0	0	0	0	0	0	3	7.67	27.0	7.25	6.00	450
23	0	9	7	1	1	1	1	0	0	5	19.67	22.0	4.75	5.00	500
24	1	2	2	0	0	0	0	0	0	4	6.67	53.0	1.00	5.00	450
25	1	6	6	1	1	1	1	0	1	7	14.33	27.0	9.75	8.00	520
26	1	5	3	1	0	0	0	0	0	5	10.67	31.0	7.25	7.00	490

หากเราตัดสินใจเลือก Hierarchical method หมายความว่า แทนที่จะให้คอมพิวเตอร์หรือโปรแกรมสำเร็จรูปเลือกตัวแปรอิสระที่เป็น Covariates ตามเกณฑ์Likelihood ratio หรือ Wald Statistic หรือเกณฑ์อื่นที่เรากำหนดเป็นเงื่อนไข เราจะเลือกลำดับตัวแปรอิสระที่จะนำเข้ามาตามดุลยพินิจของเรา มีได้ปล่อยให้มันเป็นหน้าที่ของโปรแกรมสำเร็จรูปที่จะเลือกลำดับตัวแปรอิสระแทนเรา

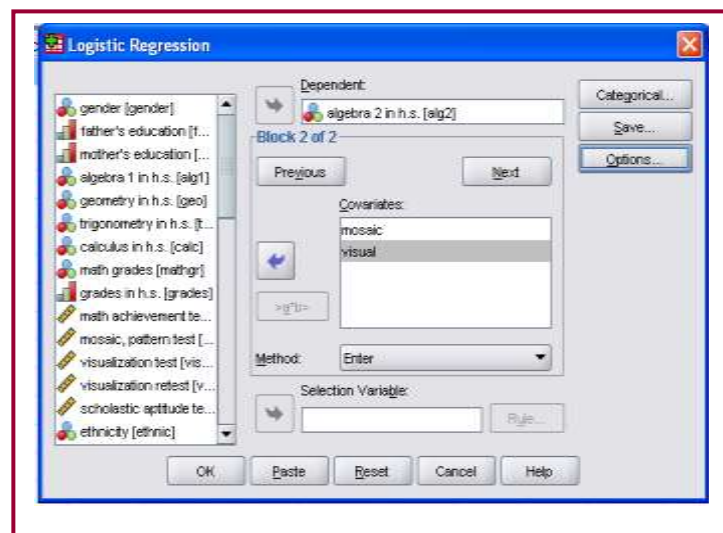
- คลิก Analyze>Regression>Binary Logistic ตามตัวอย่างภาพด้านล่าง



●คลิกเลือก algebra2 เข้าไปอยู่ในช่อง dependent สมมติว่าเราจะนำเพศ และ ระดับการศึกษาของบิดาเข้าเป็นตัวแปรอิสระที่จะใช้ในการอธิบายการตัดสินใจเลือก algebra2



จากนั้นคลิกปุ่ม Next เมนูจะเปลี่ยนไปเป็น block 2 of 2 ดังภาพถัดไป



•คลิกเลือกคะแนน mosaic test และ visualization test เข้าไปเป็นตัวแปรอิสระในขั้นตอนต่อไป แล้วคลิก OK

ต่อไปนี้เป็น computer printout ที่ได้จากโปรแกรมสำเร็จรูป

ส่วนที่หนึ่ง: แสดงข้อมูลเบื้องต้น

**Case Processing Summary**

Unweighted Cases <sup>a</sup>		N	Percent
Selected Cases	Included in Analysis	73	97.3
	Missing Cases	2	2.7
	Total	75	100.0
Unselected Cases		0	.0
Total		75	100.0

a. If weight is in effect, see classification table for the total number of cases.

ขนาดของกลุ่มตัวอย่างเท่ากับ 75 ในที่นี้มี 2 กรณีที่เป็น missing cases

**Dependent Variable Encoding**

Origin...	Internal Value
not taken	0
taken	1

แสดงการกำหนดรหัสโดยผู้เลือกเรียนจะใช้ รหัส 1 ส่วนผู้ที่ไม่ได้เลือกเรียนใช้รหัส 0

ส่วนที่สอง :แสดงClassification Table โดยใช้ค่า constant ในการพยากรณ์ และไม่มีตัวแปรอิสระใด ๆ เข้ามาช่วยในการพยากรณ์

**Block 0: Beginning**

Classification Table<sup>a,b</sup>

Observed		Predicted		Percentage Correct
		algebra 2 in h.s. not taken	algebra 2 in h.s. taken	
Step 0	algebra 2 in h.s. not taken	39	0	100.0
	algebra 2 in h.s. taken	34	0	.0
Overall Percentage				53.4

a. Constant is included in the model.  
b. The cutvalue is .500

ความถูกต้องของการพยากรณ์  
=53.4 %  
[(39+0)X100/(39+34+0+0)]

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)	
Step 0	Constant	-.137	.235	.342	1	.559	.872

Variables not in the Equation

	Score	df	Sig.
Step 0	Variables		
	gender	4.765	1 .029
	faed	11.330	1 .001
	Overall Statistics	13.123	2 .001

ตัวแปรอิสระที่เลือกไว้ในลำดับแรก  
อันได้แก่ เพศ(gender) และ  
ระดับการศึกษาของบิดา(faed)  
ต่างมีส่วนช่วยในการพยากรณ์  
พิจารณาจากStatistic scores

ส่วนที่สาม: แสดงผลที่เกิดจากการนำเอาตัวแปรอิสระชุดแรก (เพศและการศึกษาของบิดา) เข้ามา

**Block 1: Method =**

Omnibus Tests of Model Coefficients

	Chi-square	df	Sig.
Step 1	Step	13.848	2 .001
	Block	13.848	2 .001
	Model	13.848	2 .001

Model Summary

Step	-2 Log Likelihood	Cox & Snell R Square	Nagelkerke R Square
1	87.009 <sup>a</sup>	.173	.231

a. Estimation terminated at iteration number 4 because parameter estimates changed by less than .001.

จากการนำตัวแปรอิสระสองตัวเข้ามา  
ช่วยลดความคลาดเคลื่อนที่ไม่สามารถ  
อธิบายได้ลง และมีนัยสำคัญทางสถิติ

จากการนำตัวแปรอิสระ 2 ตัวเข้ามามี  
ผลทำให้ -2LLลดลงเหลือ 87.009 จาก  
100.857 (ไม่ได้มีแสดงในcomputer  
output) และตัวแปรทั้งสองมีส่วนช่วย  
อธิบายค่าของตัวแปรตาม ทั้งนี้  
พิจารณาจากค่าของ Cox & Snell R<sup>2</sup>  
หรือ Nagelkerke R<sup>2</sup>

Classification Table<sup>a</sup>

Observed		Predicted		Percentage Correct
		algebra 2 in h.s. not taken	algebra 2 in h.s. taken	
Step 1	algebra 2 in h.s. not taken	32	7	82.1
	algebra 2 in h.s. taken	15	19	55.9
Overall Percentage				69.9

a. The cutvalue is .500

Classification Table แสดงให้เห็นผล  
จากการนำตัวแปรอิสระ 2 ตัวเข้ามา ซึ่ง  
มีผลทำให้ความถูกต้องของการ  
พยากรณ์ดีขึ้น เป็นร้อยละ 69.86  
(32+19)X100/(32+7+15+19)

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I. for EXP(B)	
							Lower	Upper
Step 1 <sup>a</sup>								
gender	-.749	.525	2.038	1	.153	.473	.169	1.322
faed	.280	.099	8.050	1	.005	1.323	1.090	1.604
Constant	-1.050	.624	2.829	1	.093	.350		

a. Variable(s) entered on step 1: gender, faed.

เมื่อพิจารณาจาก Wald Statistic เพศ(gender) ไม่มีนัยสำคัญทางสถิติ ในขณะที่การศึกษาของบิดา(faed) มีนัยสำคัญ

ส่วนที่สี่ : แสดงผลที่เกิดจากการนำตัวแปรชุดที่สอง (mosaic test , visualization test) เข้ามา

## Block 2: Method =

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	7.741	2	.021
	Block	7.741	2	.021
	Model	21.589	4	.000

การนำตัวแปรชุดที่สองเข้ามา มีผลทำให้ -2LL ลดลงอีก 7.741 และมีนัยสำคัญทางสถิติ

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	79.267 <sup>a</sup>	.256	.342

a. Estimation terminated at iteration number 4 because parameter estimates changed by less than .001.

ขณะนี้ -2LL ลดลงจาก 87.009 ลงเหลือ 79.267 และ Cox & Snell R<sup>2</sup> และ Nagelkerke R<sup>2</sup> มีค่าสูงขึ้น

Classification Table<sup>a</sup>

		Observed		Predicted		Percentage Correct
				algebra 2 in h.s.		
		not taken	taken	not taken	taken	
Step 1	algebra 2 in h.s.	not taken		31	8	79.5
		taken		9	25	73.5
Overall Percentage						76.7

a. The cut value is .500

จากการเพิ่มตัวแปรอิสระอีก 2 ตัวเข้ามา มีผลทำให้ความถูกต้องในการพยากรณ์สูงขึ้น เป็นร้อยละ 76.7 [(31+25)X100/(31+8+9+25)]

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I. for EXP(B)	
							Lower	Upper
Step 1 <sup>a</sup>								
gender	-.428	.575	.554	1	.457	.652	.211	2.012
faed	.290	.102	8.101	1	.004	1.337	1.095	1.633
mosaic	-.034	.031	1.207	1	.272	.967	.910	1.027
visual	.185	.077	5.773	1	.016	1.203	1.035	1.398
Constant	-1.335	1.106	1.457	1	.227	.263		

a. Variable(s) entered on step 1: mosaic, visual.

ทั้งผลการศึกษาของบิดาและ Visualization test ล้วนเป็นตัวแปรสำคัญที่มีส่วนอธิบายค่าของตัวแปรตาม

ข้อสังเกต: หากผู้วิเคราะห์ได้กลับไปพิจารณา Statistics Talks#17( logistic regression workshop#1-enter method ) จะเห็นได้ว่า ผลที่ได้จากโปรแกรมสำเร็จรูปให้ค่าเหมือนกัน เมื่อเป็นเช่นนี้อาจคิดไปว่า ในเมื่อผลที่ได้เหมือนกัน ดังนั้นผู้วิเคราะห์จะใช้วิธีใดก็ได้เพราะในที่สุดแล้วจะได้ผลเหมือนกัน อย่างไรก็ตามอยากให้ข้อคิดว่า หากจำนวนตัวแปรอิสระมีมากและผู้วิเคราะห์ไม่มีข้อมูลทางทฤษฎีว่ามีตัวแปรใดบ้างที่มีผลต่อตัวแปรตาม เมื่อนั้นการเลือกใช้ Enter method เป็นวิธีวิเคราะห์ logistic regression ที่สมเหตุสมผล แต่หากตัวแปรมีน้อยและผู้วิเคราะห์พอมีข้อมูลพื้นฐานทางทฤษฎีมาบ้างว่าตัวแปรอิสระตัวใดที่มีความสัมพันธ์กับตัวแปรตาม เมื่อนั้นการเลือกใช้ Logistic regression แบบ hierarchical ก็เป็นวิธีที่น่าจะได้รับการพิจารณาใช้ก่อนเป็นอันดับแรก ที่กล่าวเช่นนี้ก็เนื่องจาก hierarchical method จะให้ข้อมูลเชิงลึกมากกว่า enter method ตัวอย่างเช่นในขั้นตอนแรกที่มีการนำตัวแปรอิสระเข้ามา ตัวแปรอิสระบางตัวมีนัยสำคัญทางสถิติ แต่เมื่อนำตัวแปรอิสระอื่นเข้ามาในขั้นตอนที่สอง ตัวแปรอิสระเหล่านั้นกลับไม่มีนัยสำคัญทางสถิติ ผู้วิเคราะห์จะสามารถชี้ชัดได้ว่า อาจเป็นเพราะในแต่ละขั้นตอนของการนำตัวแปรเข้า ตัวแปรอิสระที่นำเข้าแต่ละครั้งอาจมีความสัมพันธ์ระหว่างกัน เป็นผลให้ตัวแปรอิสระที่นำเข้าแต่แรกถูกลดความสำคัญลงไป

Contribution this issue: ดร. ดนัย ปัตตพงษ์

อยากเรียนรู้การนำสถิติข้างต้นนี้ไปใช้ในการวิจัยระดับสารนิพนธ์ (independent study) วิทยานิพนธ์ (thesis) คุษนิพนธ์(dissertation) ปรึกษาได้ที่ [dpattaphongse@gmail.com](mailto: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<sup>®</sup> สำหรับใช้ทดสอบ/เรียนรู้ศัพท์(ประกอบด้วยแบบฝึกหัดและเฉลยกว่า 90 บทครอบคลุมศัพท์ระดับ SAT/IELTS/TOEFL กว่า 12,000 คำ) และไวยากรณ์อังกฤษ (ประกอบด้วยแบบฝึกหัดและเฉลยกว่า 160 บทหรือกว่า 10,000 ข้อครอบคลุมเนื้อหาในระดับอุดมศึกษาและTOEFL) มาพร้อมกับไฟล์เสียง/ไฟล์ข้อมูล/ฯลฯ อีกมาก(กำหนด DEBUT 1 เมษายน 2569)