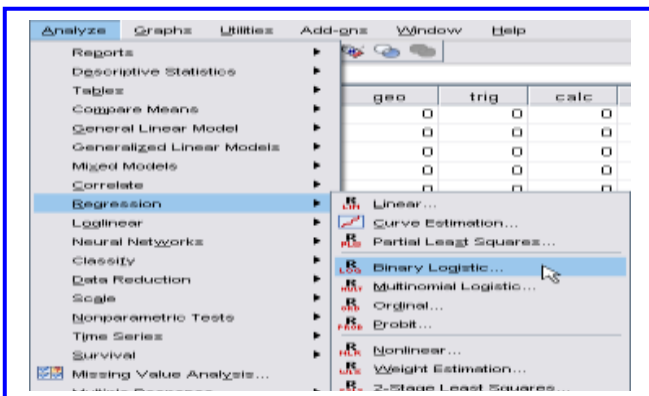


Logistic Regression Workshop #2-Stepwise method

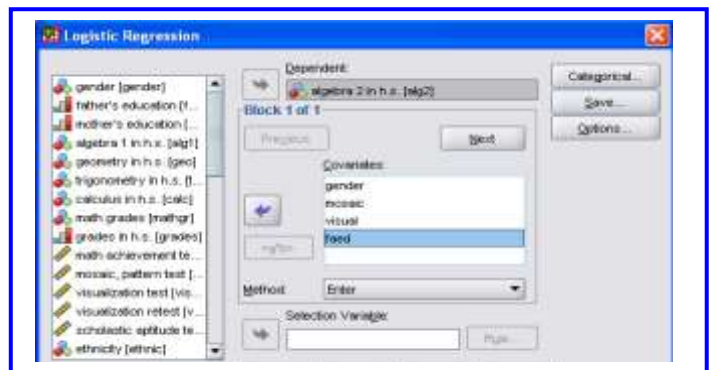
ภาพด้านล่างแสดงให้เห็นลักษณะของข้อมูลที่เราต้องการพยากรณ์การเลือกเรียนวิชาพีชคณิต2 โดยอาศัยข้อมูลจากเพศของนักเรียน ระดับการศึกษาของบิดา ผลคะแนนที่ได้จากการทดสอบวิชาว่าด้วยลวดลาย (mosaic test) และผลคะแนนที่ได้จากการทดสอบวิชาว่าด้วยการมองเห็น (visualization test)

	gender	faed	maed	alg1	alg2	geo	trig	calc	mathgr	grades	mathach	mosaic	visual	visua02	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	56.0	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	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	250
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	0	6	21.00	27.0	11.00	9.50	500
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	25.0	4.75	4.00	400
16	0		3	1	1	1	0	0	1	7	19.67	33.0	13.50	9.50	590
17	0	3	3	1	0	1	0	0	0	7	7.67	33.0	-0.25	1.00	430
18	0	9	7	1	0	1	0	0	0	5	14.33	27.0	6.00	5.00	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	25.0	8.75	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

- คลิก Analyze>Regression>Binary Logistic



- • คลิกเลือก alg2 ไปไว้ในช่องตัวแปรตาม คลิกเลือกเพศ(gender) การศึกษาของบิดา(faed) visualization test และ mosaic test เข้าไปไว้ในช่องตัวแปรอิสระ(Covariates)



●คลิกที่ปุ่ม Enter จะแสดงวิธีการนำตัวแปรเข้าหลายวิธีดังนี้



(1) Enter:จะนำตัวแปรทั้งหมดเข้ามาในคราวเดียวพร้อมกันทั้งหมด

(statistics talks #17-logistic regression workshop #1ใช้วิธีนี้)

(2) Forward Selection(Conditional) จะทำการคัดเลือกโดย

อัตโนมัติทีละขั้นทีละตอน มีการนำเอาตัวแปรอิสระเข้ามา

ตามลำดับนัยสำคัญของscore statistic และจะทำการถอนตัว

แปรอิสระออกตามprobability ของlikelihood-ratio statistic

(3) Forward Selection(Likelihood Ratio) จะทำการคัดเลือกโดยอัตโนมัติทีละขั้นทีละตอน มีการนำเอาตัวแปรอิสระเข้ามาตามลำดับนัยสำคัญของscore statistic และจะทำการถอนตัวแปรอิสระออกโดยพิจารณาจากค่าmaximum partial likelihood estimates

(4) Forward Selection (Wald) จะทำการคัดเลือกโดยอัตโนมัติทีละขั้นทีละตอน มีการนำเอาตัวแปรอิสระเข้ามาตามลำดับนัยสำคัญของscore statistic และจะทำการถอนตัวแปรอิสระออกโดยพิจารณาจาก Probability (Wald statistic)ที่คำนวณได้

(5) Backward Elimination (Conditional) จะทำการคัดเลือกตัวแปรอิสระออกโดยอัตโนมัติเป็นขั้นเป็นตอน โดยคำนวณจาก Probability ของ likelihood-ratio statistic

(6) Backward Elimination (Likelihood Ratio) จะทำการคัดเลือกตัวแปรอิสระออกโดยอัตโนมัติเป็นขั้นเป็นตอน โดยคำนวณจาก Probability ของ likelihood-ratio statistic โดยพิจารณาจากค่า maximum partial likelihood estimates

(7) Backward Elimination (Wald) จะทำการคัดเลือกตัวแปรอิสระออกโดยอัตโนมัติเป็นขั้นเป็นตอน โดยคำนวณจาก Probability ของ Wald statistic ที่คำนวณได้

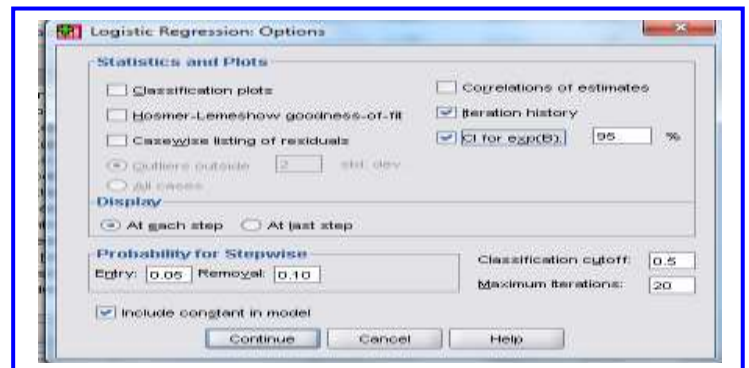
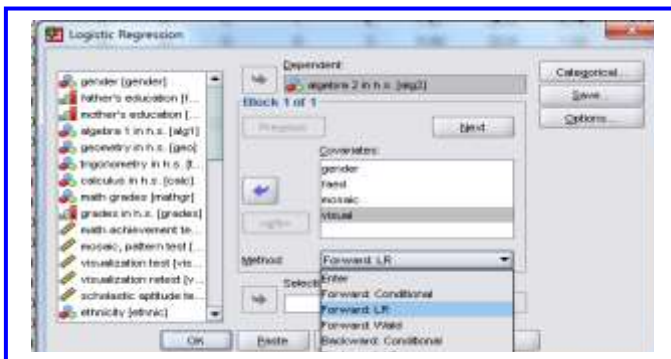
ในเอกสารวิชาการฉบับนี้เราจะมาเปรียบเทียบผลที่ได้จากการทำ Logistic regression โดยใช้วิธี (3) (4) (6) (7)

Forward Selection (Likelihood Ratio)

●คลิกเลือกForward:LR

→ ●คลิกปุ่ม Optionsในส่วนStatistics and Plots คลิกเลือกIteration

history และ CI (Confidence interval) for exp(B)



ตารางต่อไปนี้แสดง computer output

Case Processing Summary

Unweighted Cases ^a		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.

Dependent Variable Encoding

Origin	Internal Value
not taken	0
taken	1

แสดงให้เห็นขนาดของกลุ่มตัวอย่าง และ
จำนวนกรณีที่เป็น missing values
ตลอดจนการกำหนดรหัสสำหรับตัวแปรตาม

Block 0: Beginning

Iteration History^{a,b,c}

Iteration		-2 Log likelihood	Coefficients	
			Constant	
Step 0	1	100.857	-.137	
	2	100.857	-.137	

a. Constant is included in the model.
b. Initial -2 Log Likelihood: 100.857
c. Estimation terminated at iteration number 2 because parameter estimates changed by less than .001.

หากไม่ใช้ตัวแปรอิสระมาช่วยในการพยากรณ์
ใช้แต่เฉพาะข้อมูลที่มีอยู่ ความคลาดเคลื่อน
ระหว่างค่าที่สังเกตเห็นและค่าที่พยากรณ์ได้ที่ไม่
สามารถอธิบายได้เท่ากับ 100.857

Classification Table^{a,b}

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

a. Constant is included in the model.
b. The cut value is .500

ความสามารถในการพยากรณ์มีความ
ถูกต้องร้อยละ 53.4
 $[(39 \times 100) / (39 + 34)] = 53.42$

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
mosaic	.786	1	.375
visual	8.842	1	.003
Overall Statistics	19.771	4	.001

เป็นการยืนยันว่า การพยากรณ์ไม่ได้ใช้ประโยชน์
จากตัวแปรอิสระเลยแม้แต่ตัวเดียว แต่จะเห็นได้
จากค่า score statistic และค่า sig. ว่า การศึกษา
ของบิดา(faed) และ visualization test
เป็นตัวแปรอิสระที่ควรนำเข้ามาช่วยในการ
พยากรณ์

Block 1: Method = Forward Stepwise (Likelihood Ratio)

Iteration History^{a,b,c,d}

Iteration		-2 Log likelihood	Coefficients		
			Constant	faed	visual
Step 1	1	89.136	-1.459	.280	
	2	89.047	-1.583	.307	
	3	89.047	-1.585	.308	
	4	89.047	-1.586	.308	
Step 2	1	81.630	-2.176	.265	.161
	2	81.213	-2.559	.297	.193
	3	81.212	-2.586	.300	.196
	4	81.212	-2.586	.300	.196

a. Method: Forward Stepwise (Likelihood Ratio)
b. Constant is included in the model.
c. Initial -2 Log Likelihood: 100.857
d. Estimation terminated at iteration number 4 because parameter estimates changed by less than .001.

มีการ simulate โดยนำการศึกษาของบิดาเข้ามา
ก่อน ต่อเมื่อค่า parameter ที่ใช้ประเมินค่อนข้าง
คงที่ไม่เปลี่ยนแปลงแล้ว จึงทำการ simulate โดย
นำเอา visualization test เข้ามา โปรดสังเกตความ
คลาดเคลื่อนที่ไม่สามารถอธิบายได้จะลดลงเป็น
ลำดับ

Omnibus Tests of Model Coefficients

Step	Step	Chi-Square	df	Sig.
Step 1	Step	11.910	1	.001
	Block	11.910	1	.001
Step 2	Step	7.925	1	.005
	Block	19.645	2	.000

Model Summary

Step	-2 Log Likelihood	Cox & Snell R Squares	Nagelkerke R Squares
1	99.047 ^a	.149	.199
2	91.122 ^a	.230	.319

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

Classification Table^a

	Observed	Predicted	Iteration 2 in h.s.		Percentage Correct
			not taken	taken	
Step 1	algebra 2 in h.s.	not taken	20	6	64.6
	taken	not taken	10	19	62.9
	Overall Percentage				69.9
Step 2	algebra 2 in h.s.	not taken	20	9	76.9
	taken	not taken	10	24	70.6
	Overall Percentage				74.0

a. The cutvalue is .500.

จากการนำเอาการศึกษาของบิดา(faed) เข้ามา มีส่วนลดความคลาดเคลื่อนลงเท่ากับค่าสถิติไคร์สแควร์(100.857-89.047)ซึ่งมีนัยสำคัญ และเมื่อนำเอา visualization test เข้ามา ก็ยังทำให้ความคลาดเคลื่อนลดลงอีก 7.835(89.047-81.212) รวมผลจากตัวแปรอิสระสองตัวทำให้ความคลาดเคลื่อนลดลง 19.645 และความถูกต้องในการพยากรณ์เท่ากับร้อยละ 74.0

Variables in the Equation

Step	Variable	B	SE	Wald	df	Sig.	Exp(B)	95.0% C.I. for Exp(B)	Lower	Upper
Step 1	faed	.268	.097	10.108	1	.001	1.301	1.125	1.488	
	Constant	-1.595	.512	9.592	1	.002	.205			
Step 2	faed	.200	.098	9.135	1	.003	1.250	1.111	1.409	
	visual	.198	.074	8.918	1	.009	1.216	1.091	1.407	
	Constant	-2.208	.808	14.212	1	.000	.279			

a. Variable(s) entered on step 1: faed.
b. Variable(s) entered on step 2: visual.

Model & Term Homestead

Variable	Model Log Likelihood	Change in -2 Log Likelihood	df	Sig. of the Change
Step 1: faed	-50.428	11.910	1	.001
Step 2: faed, visual	-45.703	10.375	1	.001
	-44.522	7.925	1	.005

Variables not in the Equation

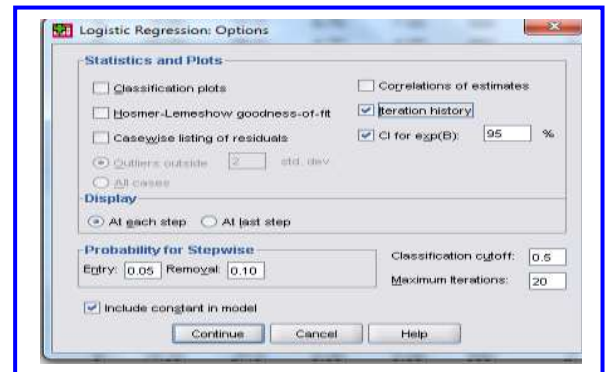
Step	Variable	B	SE	df	Sig.
Step 1	gender	2.072		1	.150
	moses	1.232		1	.249
	Overall Statistics	7.621		2	.020
Step 2	gender	6.776		1	.011
	moses	1.380		1	.241
	Overall Statistics	1.896		2	.389

จะเห็นได้ว่า ตามวิธีนี้ไม่นำเอาตัวแปรอิสระอีกสองตัว (เพศและmosaic test) เข้ามาช่วยในการพยากรณ์เลย มีผลทำให้ความแม่นยำในการพยากรณ์ลดลงเล็กน้อย เมื่อเปรียบเทียบกับวิธี Entry method (Statistics talks #17) หรือ Hierarchical method(Statistics talks#19)

Forward Selection (Wald)

- คลิกเลือก Forward:Wald

- คลิกปุ่มoptions ในส่วนที่เป็นStatistics and Plots
คลิกเลือกiteration history และ CI for exp(B)



ตารางต่อไปนี้แสดง computer output

Case Processing Summary

Unweighted Cases ^a		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.

Dependent Variable Encoding

Origin	Internal Value
not taken	0
taken	1

แสดงให้เห็นขนาดของกลุ่มตัวอย่าง และจำนวนกรณีที่เป็น missing values ตลอดจนการกำหนดรหัสสำหรับตัวแปรตาม

Block 0: Beginning

Iteration History^{a,b}

Iteration	-2 Log Likelihood	Coefficients	
		Constant	algebra 2 in h.s.
Step 0	100.857	.137	
1	89.057		
2	89.057		

- a. Constant is included in the model.
- b. Initial -2 Log Likelihood: 100.857
- c. Estimation terminated at iteration number 2 because parameter estimates changed by less than .001.

Classification Table^a

Observed	algebra 2 in h.s.	Predicted		Percentage Correct
		not taken	taken	
Step 0	algebra 2 in h.s.	not taken	taken	
		38	0	100.0
		34	0	.0
Overall Percentage				53.4

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

Variables in the Equation

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

Variables not in the Equation

Step 0	Variables	Score	df	Sig.
	gender	4.765	1	.029
	faed	11.230	1	.001
	mosaic	.798	1	.375
	visual	8.842	1	.003
Overall Statistics		19.771	4	.001

Block 1: Method = Forward Stepwise (Wald)

Iteration History^{a,b,c}

Iteration	-2 Log Likelihood	Coefficients			
		Constant	faed	visual	gender
Step 1	1	89.136	-1.459	.300	
	2	89.047	-1.583	.307	
	3	89.047	-1.585	.300	
	4	89.047	-1.585	.300	
Step 2	1	81.630	-2.176	.255	.161
	2	81.212	-2.588	.287	.193
	3	81.212	-2.588	.300	.196
	4	81.212	-2.588	.300	.196

- a. Method: Forward Stepwise (Wald)
- b. Constant is included in the model.
- c. Initial -2 Log Likelihood: 100.857
- d. Estimation terminated at iteration number 4 because parameter estimates changed by less than .001.

Overall Tests of Model Coefficients

Step	Chi-square	df	Sig.
Step 1			
Step	11.810	1	.001
Block	11.810	1	.001
Model	11.810	1	.001
Step 2			
Step	7.835	1	.005
Block	19.645	2	.000
Model	19.645	2	.000

Model Summary

Step	-2 Log Likelihood	Cox & Snell R-Square	Nagelkerke R-Square
1	89.047*	.149	.199
2	81.212*	.238	.315

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

Classification Table^a

Observed	algebra 2 in h.s.	not taken	Predicted		Percentage Correct
			not taken	taken	
Step 1	algebra 2 in h.s.	not taken	33	6	84.6
		taken	16	18	52.9
Overall Percentage					89.9
Step 2	algebra 2 in h.s.	not taken	30	9	76.9
		taken	10	24	70.6
Overall Percentage					74.0

- a. The cutvalue is: .500

หากไม่ใช้ตัวแปรอิสระมาช่วยในการพยากรณ์

ใช้แต่เฉพาะข้อมูลที่มีอยู่ ความคลาดเคลื่อนระหว่างค่าที่สังเกตเห็นและค่าที่พยากรณ์ได้ที่ไม่สามารถอธิบายได้เท่ากับ 100.857ความสามารถในการพยากรณ์มีความถูกต้องร้อยละ 53.4 $[(39 \times 100) / (39 + 34)] = 53.42$

เป็นการยืนยันว่า การพยากรณ์ไม่ได้ใช้ประโยชน์จากตัวแปรอิสระเลยแม้แต่นิดเดียว แต่จะเห็นได้จากค่า score statistic และค่า sig. ว่า การศึกษาของบิดา (faed) และ visualization test เป็นตัวแปรอิสระที่ควรนำมาช่วยในการพยากรณ์

มีการ simulate โดยนำการศึกษาของบิดาเข้ามาก่อน ต่อเมื่อค่าparameter ที่ใช้ประเมินค่อนข้างคงที่ไม่เปลี่ยนแปลงแล้วจึงทำการ simulate โดยนำเอา visualization test เข้ามาโปรดสังเกตความคลาดเคลื่อนที่ไม่สามารถอธิบายได้จะลดลงเป็นลำดับ

จากการนำเอาการศึกษาของบิดา(faed) เข้ามา มีส่วนลดความคลาดเคลื่อนลงเท่ากับค่าสถิติไคร์สแควร์ (100.857-89.047)ซึ่งมีนัยสำคัญ และเมื่อนำเอา visualization test เข้ามา ก็ยิ่งทำให้ความคลาดเคลื่อนลดลงอีก 7.835 (89.047-81.212) รวมผลจากตัวแปรอิสระสองตัวทำให้ความคลาดเคลื่อนลดลง 19.645 และความถูกต้องในการพยากรณ์เท่ากับร้อยละ 74.0

Variables in the Equation									
	B	S.E.	Wald	df	Sig.	Exp(B)	Lower	Upper	
Step 1 ^a	faed	.308	.897	10.106	1	.001	1.361	1.125	1.648
	Constant	-1.585	.511	9.592	1	.002	.205		
Step 2 ^a	faed	.300	.898	9.135	1	.003	1.350	1.111	1.638
	visual	.198	.874	6.919	1	.009	1.216	1.051	1.407
	Constant	-2.586	.886	14.233	1	.000	.075		

a. Variable(s) entered on step 1: faed.
b. Variable(s) entered on step 2: visual.

Variables not in the Equation			
Step	Variables	Score	Sig.
Step 1	gender	2.072	1
	mosaic	1.322	1
	visual	7.681	1
	Overall Statistics	9.454	3
Step 2	gender	.876	1
	mosaic	1.380	1
	Overall Statistics	1.905	2

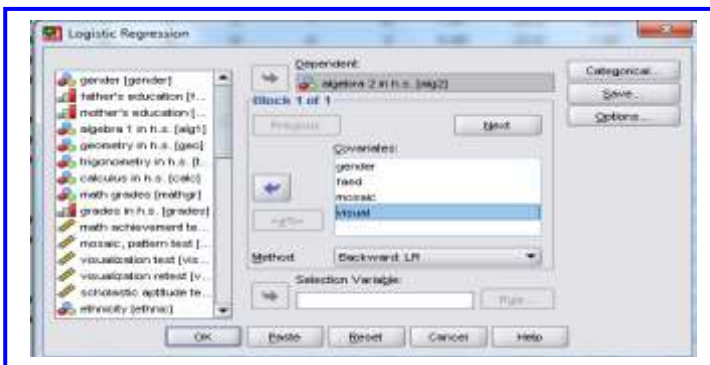
จะเห็นได้ว่า ตามวิธีนี้ไม่นำเอาตัวแปรอิสระอีกสองตัว(เพศและmosaic test) เข้ามาช่วยในการพยากรณ์เลยมีผลทำให้ความแม่นยำในการพยากรณ์ลดลงเล็กน้อยเมื่อเปรียบเทียบกับวิธี Entry method (Statistics talks #17) หรือ Hierarchical method(Statistics talks#19)

Backward Elimination: LR

- คลิกเลือก Backward:LR

→ ●คลิกปุ่มoptions ในส่วนที่เป็นStatistics and Plots

คลิกเลือกiteration history และ CI for exp(B)



ต่อไปนี้เป็น computer output

Case Processing Summary			
Unweighted Cases ^a		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.

Dependent Variable Encoding	
Origin	Internal Value
not taken	0
taken	1

แสดงให้เห็นขนาดของกลุ่มตัวอย่าง และจำนวนกรณีที่เป็น missing values ตลอดจนการกำหนดรหัสสำหรับตัวแปรตาม

Block 0: Beginning				
Classification Table ^{a,b}				
Observed	Predicted	algebra 2 in h.s.		Percentage Correct
		not taken	taken	
Step 0 algebra 2 in h.s. not taken		39	0	100.0
taken		34	0	0
Overall Percentage				53.4

a. Constant is included in the model.
b. The cut value is: .500

Variables in the Equation						
	B	S.E.	Wald	df	Sig.	Exp(B)
Step 0 Constant	-.137	.235	34.2	1	.559	0.72

Variables not in the Equation			
Step	Variables	Score	Sig.
Step 0	gender	4.766	1
	faed	11.330	1
	mosaic	.756	1
	visual	6.642	1
Overall Statistics		19.777	4

หากไม่ใช้ตัวแปรอิสระมาช่วยในการพยากรณ์ความสามารถในการพยากรณ์มีความถูกต้องร้อยละ 53.4 $[(39 \times 100) / (39 + 34)] = 53.42$ ในที่นี้ค่าStatistic score ของเพศ(gender) การศึกษาของบิดา(faed) และคะแนนvisualization test มีนัยสำคัญทางสถิติ

Block 1: Method = Backward Stepwise (Likelihood Ratio)

Omnibus Tests of Model Coefficients

Step	Step	Chi-square	df	Sig.
Step 1 ^a	Step	21.589	4	.000
	Block	21.589	4	.000
	Model	21.589	4	.000
Step 2 ^a	Step	.546	1	.460
	Block	21.043	3	.000
	Model	21.043	3	.000
Step 3 ^a	Step	1.398	1	.237
	Block	19.645	2	.000
	Model	19.645	2	.000

a. A negative Chi-square value indicates that the Chi-squares value has decreased from the previous step.

Model Summary

Step	-2 Log Likelihood	Cor. & Snell R Square	Nagelkerke R Square
1	79.367 ^a	.356	.342
2	79.014 ^a	.350	.334
3	81.312 ^a	.226	.215

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

ตัวแปรอิสระทั้งหมดสามารถช่วยลด log-likelihood ได้ 21.589 หากนำตัวแปรอิสระอันดับแรกออก มีผลทำให้ log-likelihood เพิ่มขึ้น .546 และหากนำตัวแปรอิสระอันดับสองออก มีผลทำให้ log-likelihood เพิ่มขึ้น 1.398

Classification Table^a

Observed	Predicted	algebra 2 in h.s.		Percentage Correct
		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
Step 2 algebra 2 in h.s.	not taken	31	8	79.5
	taken	9	25	73.5
Overall Percentage				76.7
Step 3 algebra 2 in h.s.	not taken	30	9	76.0
	taken	10	24	70.6
Overall Percentage				74.0

a. The cut value is .500.

หลังจากนำตัวแปรอิสระทั้งสองออกแล้ว ความแม่นยำในการพยากรณ์ลดลงจาก 76.7 เหลือ 74.0

Variables in the Equation

Step	Variable	B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a	gender	-.428	.575	.554	1	.457	.652
	faed	.290	.102	8.101	1	.004	1.337
	mosaic	-.034	.031	1.207	1	.272	.967
	visual	.185	.077	5.773	1	.016	1.203
	Constant	-1.335	1.106	1.457	1	.227	.263
Step 2 ^a	faed	.307	.100	9.487	1	.002	1.360
	mosaic	-.035	.031	1.311	1	.252	.965
	visual	.199	.076	6.914	1	.009	1.220
Step 3 ^a	Constant	-1.690	1.001	2.852	1	.091	.184
	faed	.300	.099	9.135	1	.003	1.350
	visual	.196	.074	6.919	1	.009	1.216
	Constant	-2.586	.686	14.233	1	.000	.075

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

แสดงให้เห็นว่าเพศเป็นตัวแปรอิสระที่ถูกตัดออกเป็นอันดับแรก ตามด้วยคะแนน mosaic test

Model if Term Removed

Variable	Model Log Likelihood	Change in -2 Log Likelihood	df	Sig. of the Change	
Step 1	gender	-.39.907	.546	1	.460
	faed	-.44.123	8.979	1	.003
	mosaic	-.40.274	1.281	1	.258
	visual	-.42.884	6.500	1	.011
Step 2	faed	-.45.324	10.833	1	.001
	mosaic	-.40.608	1.398	1	.237
	visual	-.43.844	7.875	1	.005
Step 3	faed	-.45.793	10.375	1	.001
	visual	-.44.523	7.935	1	.005

Variables not in the Equation

Step	Variables	Score	df	Sig.
Step 2 ^a	gender	.567	1	.455
	Overall Statistics	.557	1	.455
Step 3 ^a	gender	.676	1	.411
	mosaic	1.360	1	.243
	Overall Statistics	1.905	2	.386

a. Variable(s) removed on step 2: gender.

b. Variable(s) removed on step 3: mosaic.

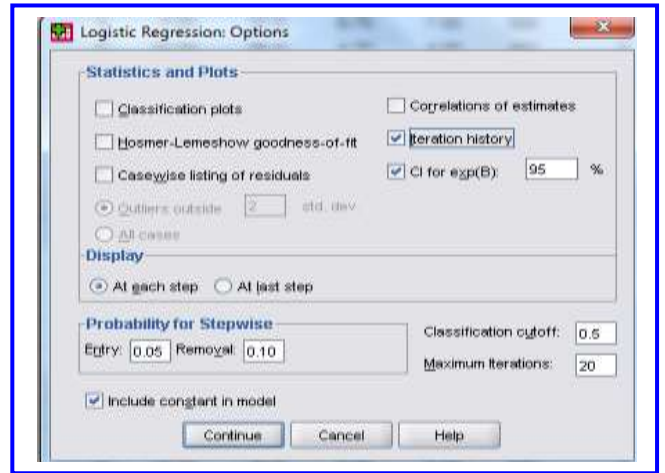
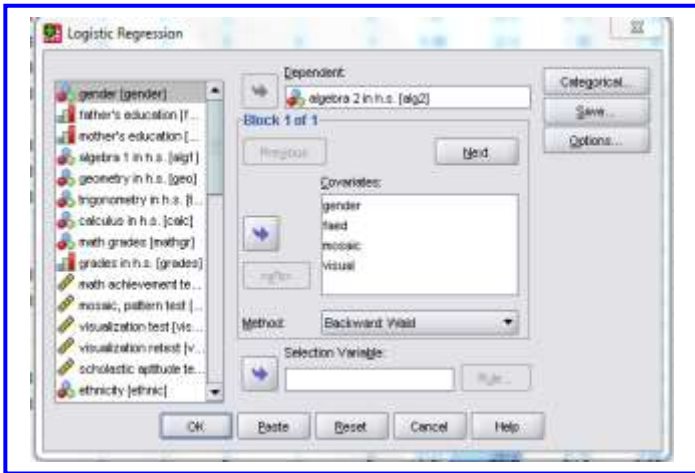
แสดงให้เห็นว่าเพศเป็นตัวแปรอิสระที่มีผลทำให้เกิดการเปลี่ยนแปลงในค่า -2LL น้อยที่สุด เมื่อตัดเพศออกไปแล้ว คะแนน mosaic test เป็นตัวแปรอิสระที่มีผลทำให้เกิดการเปลี่ยนแปลงใน -2LL น้อยที่สุดในอันดับต่อไป

Backward Elimination: Wald Statistic

- คลิกที่ Backward: Wald



- คลิกปุ่มoptions ในส่วนที่เป็นStatistics and Plots
คลิกเลือกiteration history และ CI for exp(B)



ต่อไปนี้เป็น computer output

[DataSet1] G:\hsbdataB.sav

Case Processing Summary

Unweighted Cases ^a		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.

Dependent Variable Encoding

Origin...	Internal Value
not taken	0
taken	1

แสดงให้เห็นขนาดของกลุ่มตัวอย่าง และจำนวนกรณีที่เป็น missing values ตลอดจนการกำหนดรหัสสำหรับตัวแปรตาม

Block 0: Beginning

Iteration History^{a,b,c}

Iteration	-2 Log Likelihood	Coefficients
Step 0	100.857	Constant
2	100.857	-137

a. Constant is included in the model.
b. Initial -2 Log Likelihood: 100.857
c. Estimation terminated at iteration number 2 because parameter estimates changed by less than .001.

Classification Table^{a,b}

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

a. Constant is included in the model.
b. The cut value is .500

หากไม่ใช้ตัวแปรอิสระมาช่วยในการพยากรณ์ ความสามารถในการพยากรณ์มีความถูกต้องร้อยละ 53.4 $[(39 \times 100) / (39 + 34)] = 53.42$

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)
Step 0 Constant	-.137	.235	.342	1	.558	.872

variables not in the equation

	Score	df	Sig.
Step 0 Variables gender	4.765	1	.029
faed	11.330	1	.001
mosaic	.786	1	.375
visual	8.842	1	.003
Overall Statistics	19.771	4	.001

เพศ(gender) การศึกษาของบิดา(faed) และคะแนน visualization test เป็นตัวแปรที่มีนัยสำคัญทางสถิติ ส่วนคะแนน mosaic test ไม่มีความสำคัญ

Block 1: Method = Backward Stepwise (Wald)

Iteration History^{a, b, c, d}

Iteration	Step	-2 Log Likelihood	Coefficients				
			Constant	gender	faed	mosaic	visual
Step 1	1	79.807	-1.184	-.356	.245	-.025	.151
	2	79.271	-1.328	-.423	.287	-.032	.182
	3	79.267	-1.335	-.428	.290	-.034	.185
	4	79.267	-1.335	-.428	.290	-.034	.185
Step 2	1	80.346	-1.481		.260	-.028	.181
	2	79.817	-1.680		.304	-.034	.195
	3	79.814	-1.690		.307	-.035	.199
	4	79.814	-1.690		.307	-.035	.199
Step 3	1	81.630	-2.176		.255		.161
	2	81.213	-2.558		.287		.183
	3	81.212	-2.586		.300		.196
	4	81.212	-2.586		.300		.196

- a. Method: Backward Stepwise (Wald)
- b. Constant is included in the model
- c. Initial -2 Log Likelihood: 100.857
- d. Estimation terminated at iteration number 4 because parameter estimates changed by less than .001.

มีการนำเอาเพศออกเป็นอันดับแรก ตามด้วย mosaic test

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	21.589	4	.000
	Block	21.589	4	.000
	Model	21.589	4	.000
Step 2 ^a	Step	.546	1	.460
	Block	21.043	3	.000
	Model	21.043	3	.000
Step 3 ^a	Step	-1.398	1	.237
	Block	19.645	2	.000
	Model	19.645	2	.000

- a. A negative Chi-squares value indicates that the Chi-squares value has decreased from the previous step.

Model Summary

Step	-2 Log Likelihood	Cox & Snell R Square	Nagelkerke R Square
1	79.267 ^a	.256	.342
2	79.814 ^a	.250	.334
3	81.212 ^a	.236	.315

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

แสดงผลที่เกิดกับ -2LL เมื่อมีการนำตัวแปรอิสระแต่ละตัวออก

Classification Table^a

	Observed		Predicted		Percentage Correct
			algebra 2 in h.s.		
			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
Step 2	algebra 2 in h.s.	not taken	31	8	79.5
		taken	9	25	73.5
	Overall Percentage				76.7
Step 3	algebra 2 in h.s.	not taken	30	9	76.9
		taken	10	24	70.6
	Overall Percentage				74.0

- a. The cut value is .500

จากการนำตัวแปรอิสระออกมีผลทำให้ความถูกต้องในการพยากรณ์ลดลง

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP(B)	
								Lower	Upper
Step 1 ^a	gender	-.428	.575	.954	1	.457	.652	.211	2.612
	feet	.288	.102	0.101	1	.004	1.337	1.095	1.633
	mosaic	-.014	.031	1.207	1	.272	.987	.810	1.227
	visual	.185	.077	5.773	1	.016	1.203	1.035	1.388
	Constant	-1.335	1.108	1.457	1	.227	.263		
Step 2 ^a	feet	.267	.100	9.487	1	.002	1.309	1.110	1.652
	mosaic	-.015	.031	1.311	1	.252	.985	.809	1.225
	visual	.188	.076	6.914	1	.009	1.228	1.052	1.414
	Constant	-1.658	1.001	2.852	1	.091	.184		
Step 3 ^a	feet	.308	.099	9.195	1	.003	1.359	1.111	1.639
	visual	.188	.074	6.919	1	.009	1.218	1.051	1.407
	Constant	-2.588	.686	14.232	1	.000	.075		

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

แสดงให้เห็นค่า Wald statistic ของแต่ละตัวแปรอิสระ และตัวแปรอิสระที่มี Wald statistic ต่ำที่สุดจะเป็นตัวแปรอิสระตัวแรกที่ต้องถูกตัดออก

Step	Variables	Score	df	Sig.
Step 2 ^a	Variables: gender	.557	1	.455
	Overall Statistics	.557	1	.455
Step 3 ^a	Variables: gender	.676	1	.411
	mosaic	1.360	1	.243
	Overall Statistics	1.905	2	.386

a. Variable(s) removed on step 2: gender.
b. Variable(s) removed on step 3: mosaic.

ซึ่งผลที่ได้จากการวิเคราะห์โดยวิธี Forward Selection(Likelihood Ratio) Forward Selection (Wald) Backward Elimination (Likelihood Ratio) Backward Elimination (Wald) ให้คำตอบเกี่ยวกับผลทางสถิติตรงกัน

Contribution this issue: ดร. ดนัย บัณฑิตพงศ์

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)