

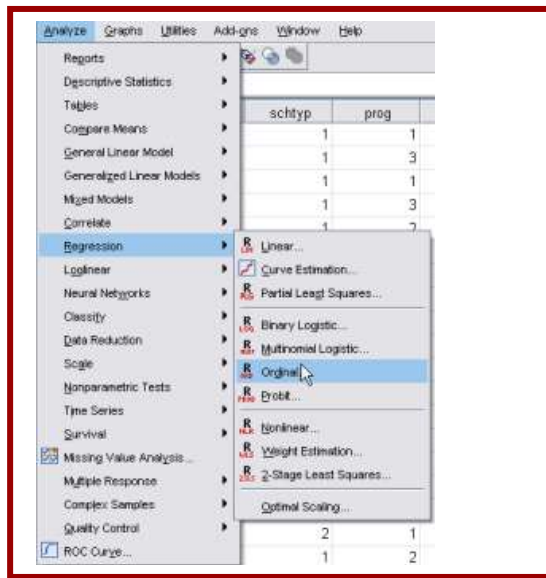


Ordered Logistic Regression Workshop #1-Drop-down menu

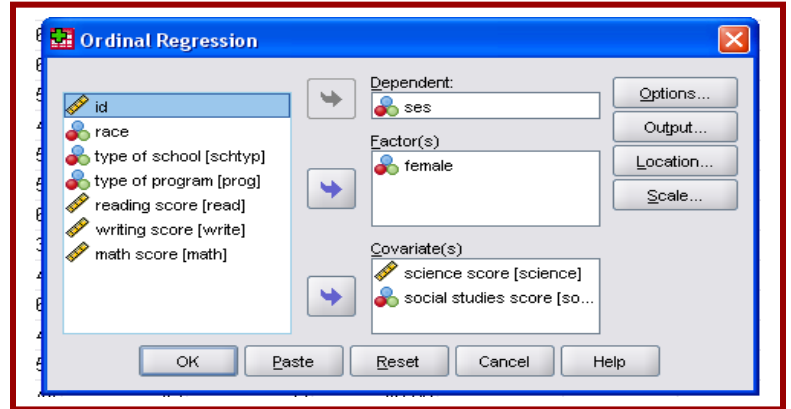
ตารางด้านล่างแสดงข้อมูลที่เราต้องการวิเคราะห์โดยใช้ Ordered logistic regression (OLOGIT) โดยตัวแปรตามคือ ses (social economic status ซึ่งมีค่า=1,2,3) ส่วนตัวแปรอิสระประกอบด้วย female(0=male,1=female) science(คะแนนสาขาวิทยาศาสตร์) socst (คะแนนวิชาสังคม)

	id	female	race	ses	schtyp	prog	read	write	math	science	socst
1	70	0	4	1	1	1	57	52	41	47	57.00
2	121	1	4	2	1	3	68	59	53	63	61.00
3	86	0	4	3	1	1	44	33	54	58	31.00
4	141	0	4	3	1	3	63	44	47	53	56.00
5	172	0	4	2	1	2	47	52	57	53	61.00
6	113	0	4	2	1	2	44	52	51	63	61.00
7	50	0	3	2	1	1	50	59	42	53	61.00
8	11	0	1	2	1	2	34	46	45	39	36.00
9	84	0	4	2	1	1	63	57	54	58	51.00
10	48	0	3	2	1	2	57	55	52	50	51.00
11	75	0	4	2	1	3	60	46	51	53	61.00
12	60	0	4	2	1	2	57	65	51	63	61.00
13	95	0	4	3	1	2	73	60	71	61	71.00
14	104	0	4	3	1	2	54	63	57	55	46.00
15	38	0	3	1	1	2	45	57	50	31	56.00
16	115	0	4	1	1	1	42	49	43	50	56.00
17	76	0	4	3	1	2	47	52	51	50	56.00
18	195	0	4	2	2	1	57	57	60	58	56.00
19	114	0	4	3	1	2	68	65	62	55	61.00
20	85	0	4	2	1	1	55	39	57	53	46.00
21	167	0	4	2	1	1	63	49	35	66	41.00
22	143	0	4	2	1	3	63	63	75	72	66.00
23	41	0	3	2	1	2	50	40	45	55	56.00
24	20	0	1	3	1	2	60	52	57	61	61.00
25	12	0	1	2	1	3	37	44	45	39	46.00
26	53	0	3	2	1	3	34	37	46	39	31.00
27	154	0	4	3	1	2	65	65	66	61	66.00
28	178	0	4	2	2	3	47	57	57	58	46.00
29	196	0	4	3	2	2	44	38	49	39	46.00
30	29	0	2	1	1	1	52	44	49	55	41.00
31	126	0	4	2	1	1	42	31	57	47	51.00
32	103	0	4	3	1	2	76	52	64	64	61.00
33	192	0	4	3	2	2	65	67	63	66	71.00
34	150	0	4	2	1	3	42	41	57	72	31.00
35	199	0	4	3	2	2	52	59	50	61	61.00
36	144	0	4	3	1	1	60	65	58	61	66.00
37	200	0	4	2	2	2	68	54	75	66	66.00
38	80	0	4	3	1	2	65	62	68	66	66.00
39	16	0	1	1	1	3	47	31	44	36	36.00

- คลิก Analyze>Regression>Ordinal แสดงได้ดังภาพต่อไปนี้



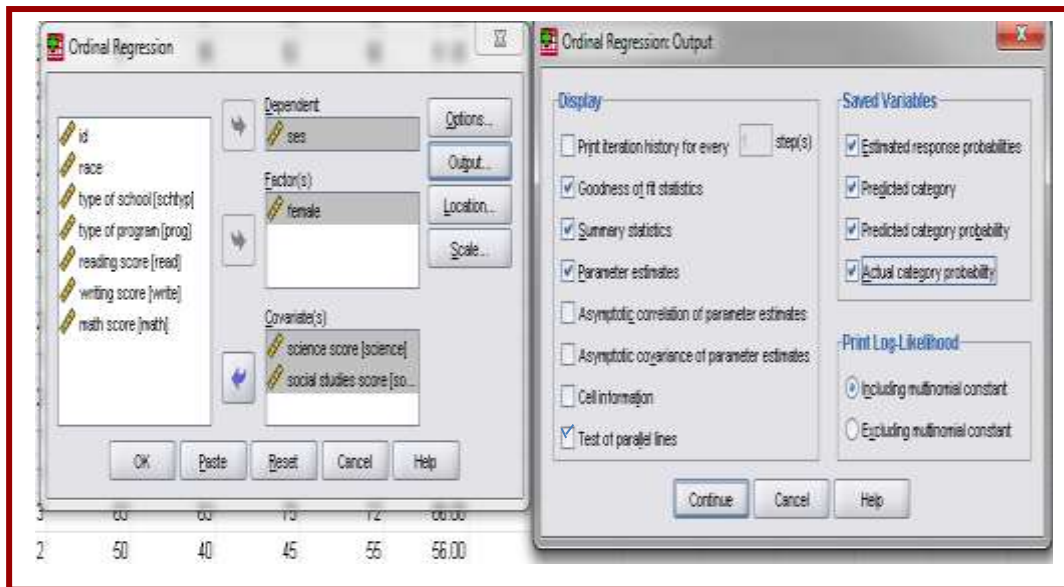
จะปรากฏ menu ดังต่อไปนี้



- คลิกเลือกตัวแปร ses ซึ่งเป็นตัวแปรตามระดับ (ordinal variable) ให้ไปอยู่ในช่องตัวแปรตาม female ให้ไปอยู่ในช่อง Factor(s) และ science / socst ให้ไปอยู่ในช่อง Covariate(s)
- คลิกปุ่ม Options ตรวจสอบให้มั่นใจว่าฟังก์ชันที่เราต้องการจะ link คือ logit ไม่ใช่ probit/Cauchit/Complementary log-log หรือ Negative log-log



- คลิกที่ปุ่ม Output ในส่วน Display เลือก Goodness of fit statistics , Summary statistics , Parameter estimates และ test of parallel lines ในส่วน Saved Variables เลือก Estimated response probabilities , Predicted category , Predicted category probability และ Actual category probability



- คลิก Continue คลิก OK

ต่อไปนี้เป็น computer output ที่ได้พร้อมทั้งความหมาย

1, Case Processing Summary

		N	Marginal Percentage
ses	low	47	23.5%
	middle	95	47.5%
	high	58	29.0%
female	male	91	45.5%
	female	109	54.5%
Valid		200	100.0%
Missing		0	
Total		200	

ขนาดของกลุ่มตัวอย่างเท่ากับ 200 เป็นหญิงมากกว่าชาย มี สถานภาพสังคมในระดับกลาง เป็นสัดส่วนที่มากที่สุด ตามด้วย สถานภาพสังคมในระดับสูง และ สถานภาพสังคมในระดับล่าง ตามลำดับ

2. Model Fitting Information

Model Fitting Information

Model	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	365.736			
Final	334.176	31.560	3	.000

Link function: Logit.

ความผันผวนในตัวแปรตามที่ไม่สามารถอธิบายได้ด้วยตัวแปรใด ๆ ใน null model (-2LL) เท่ากับ 365.736 แต่เมื่อนำตัวแปรอิสระเข้ามา ความผันผวนในตัวแปรตามที่ไม่สามารถอธิบายได้ลดลงเหลือ 334.176 ใน final model และมีนัยสำคัญทางสถิติ

3. Goodness-of-fit และ Pseudo R-Square

Goodness-of-Fit

	Chi-Square	df	Sig.
Pearson	290.757	281	.332
Deviance	289.148	281	.356

Link function: Logit.

Pseudo R-Square

Cox and Snell	.146
Nagelkerke	.166
McFadden	.075

Link function: Logit.

ผลการทดสอบ Goodness-of-fit test

พิจารณาจากค่าของ Pearson's และ Deviance Chi-square มีค่า sig.>0.05

ดังนั้นการใช้ ordered logistic regression ในการหาความสัมพันธ์ระหว่างตัวแปรตามและตัวแปรอิสระมีความเหมาะสม

อนึ่งค่า Pseudo R-square แม้ไม่สูง แต่แสดงให้เห็นประโยชน์จากการใช้ตัวแปรอิสระในการพยากรณ์ค่าของตัวแปรตาม

4. Parameter Estimates

Parameter Estimates

	Estimate	Std. Error	Wald	df	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Threshold [ses = 1]	3.237	.852	14.425	1	.000	1.567	4.908
[ses = 2]	5.588	.922	36.749	1	.000	3.781	7.395
Location science	.030	.016	3.584	1	.058	-.001	.061
socst	.053	.015	12.777	1	.000	.024	.082
[female=0]	.482	.279	3.000	1	.083	-.063	1.028
[female=1]	0 ^a	.	.	0	.	.	.

Link function: Logit.

a. This parameter is set to zero because it is redundant.

Threshold หมายถึงจุดแบ่งแยกค่าของตัวแปรตาม(ในที่นี้คือ สถานภาพทางสังคมระหว่างต่ำ กลางและสูง)ออกเป็นระดับต่าง ๆกัน ในที่นี้ thresholds มีนัยสำคัญ ดังนั้นการแยกแยะสถานภาพทางสังคม (ses) ซึ่งเป็นตัวแปรตาม ออกเป็น ต่ำ กลาง สูง โดยใช้ ordered logistic regression มีความเหมาะสมถูกต้อง และเราไม่จำเป็นต้องตัดหรือยุบรวมลำดับที่ของตัวแปรตามแต่อย่างใด

Location แสดงค่าที่ได้จากตัวแปรอิสระ ในที่นี้ regression coefficient ของตัวแปร socst มีนัยสำคัญทางสถิติ ส่วน regression coefficients ของตัวแปรอิสระอีกสองตัว (science และ female) ไม่มีนัยสำคัญทางสถิติ

5. Test of Parallel Lines

Test of Parallel Lines^a

Model	-2 Log Likelihood	Chi-Square	df	Sig.
Null Hypothesis	334.176			
General	331.987	2.189	3	.534

The null hypothesis states that the location parameters (slope coefficients) are the same across response categories.

a. Link function: Logit.

ผลจาก parallel line test ไม่มีนัยสำคัญ แสดงว่าการกำหนดให้ใช้ cumulative logit function เพียงสมการเดียวที่ค่าของ regression coefficients ของตัวแปรอิสระทุกตัวคงที่ ในขณะที่ intercept มีค่าแบ่งออกได้ตามลำดับของตัวแปรตาม หรือที่รู้จักกันว่า proportional odds assumption มีความเหมาะสม และการใช้ OLOGIT ในการหาความสัมพันธ์ระหว่างตัวแปรอิสระที่กำหนดกับตัวแปรตามที่มีค่าแบ่งออกเป็นระดับมีความน่าเชื่อถือในทางสถิติ

เราจะมาพิจารณาผลอื่น ๆ ที่ได้จากการใช้โปรแกรมสำเร็จรูป

ses	schtyp	prog	read	write	math	science	socst	EST1_1	EST2_1	EST3_1	PRE_1	PCP_1	ACP_1
1	1	1	57	52	41	47	57.00	0.16	0.50	0.34	2	0.50	0.16
2	1	3	68	59	53	63	61.00	0.13	0.48	0.39	2	0.48	0.48
3	1	1	44	33	54	58	31.00	0.35	0.50	0.15	2	0.50	0.15
3	1	3	63	44	47	53	56.00	0.14	0.49	0.37	2	0.49	0.37
2	1	2	47	52	57	53	61.00	0.11	0.46	0.43	2	0.46	0.46
2	1	2	44	52	51	63	61.00	0.08	0.41	0.51	3	0.51	0.41
2	1	1	50	59	42	53	61.00	0.11	0.46	0.43	2	0.46	0.46
2	1	2	34	46	45	39	36.00	0.42	0.46	0.12	2	0.46	0.46
2	1	1	63	57	54	58	51.00	0.15	0.50	0.34	2	0.50	0.50
2	1	2	57	55	52	50	51.00	0.19	0.52	0.29	2	0.52	0.52
2	1	3	60	46	51	53	61.00	0.11	0.46	0.43	2	0.46	0.46
2	1	2	57	65	51	63	61.00	0.08	0.41	0.51	3	0.51	0.41
3	1	2	73	60	71	61	71.00	0.05	0.32	0.62	3	0.62	0.62
3	1	2	54	63	57	55	46.00	0.21	0.53	0.27	2	0.53	0.27
1	1	2	45	57	50	31	56.00	0.24	0.53	0.23	2	0.53	0.24
1	1	1	42	49	43	50	56.00	0.15	0.50	0.36	2	0.50	0.15
3	1	2	47	52	51	50	56.00	0.15	0.50	0.36	2	0.50	0.36
2	2	1	57	57	60	58	56.00	0.12	0.47	0.40	2	0.47	0.47
3	1	2	68	65	62	55	61.00	0.11	0.45	0.45	3	0.45	0.45
2	1	1	55	39	57	53	46.00	0.22	0.53	0.26	2	0.53	0.53
2	1	1	63	49	35	66	41.00	0.20	0.52	0.28	2	0.52	0.52
2	1	3	63	63	75	72	66.00	0.05	0.31	0.64	3	0.64	0.31
2	1	2	50	40	45	55	56.00	0.13	0.48	0.38	2	0.48	0.48
3	1	2	60	52	57	61	61.00	0.09	0.42	0.49	3	0.49	0.49
2	1	3	37	44	45	39	46.00	0.30	0.52	0.18	2	0.52	0.52
2	1	3	34	37	46	39	31.00	0.48	0.42	0.09	1	0.48	0.42
3	1	2	65	65	66	61	66.00	0.07	0.37	0.56	3	0.56	0.56
2	2	3	47	57	57	58	46.00	0.19	0.52	0.29	2	0.52	0.52
3	2	2	44	38	49	39	46.00	0.30	0.52	0.18	2	0.52	0.18
1	1	1	52	44	49	55	41.00	0.25	0.53	0.22	2	0.53	0.25
2	1	1	42	31	57	47	51.00	0.20	0.52	0.27	2	0.52	0.52
3	1	2	76	52	64	64	61.00	0.08	0.40	0.51	3	0.51	0.51
3	2	2	65	67	63	66	71.00	0.05	0.30	0.66	3	0.66	0.66
2	1	3	42	41	57	72	31.00	0.26	0.53	0.21	2	0.53	0.53
3	2	2	52	59	90	61	61.00	0.09	0.42	0.49	3	0.49	0.49
3	1	1	60	65	58	61	66.00	0.07	0.37	0.56	3	0.56	0.56
2	2	2	68	54	75	66	66.00	0.06	0.34	0.60	3	0.60	0.34
3	1	2	65	62	68	66	66.00	0.06	0.34	0.60	3	0.60	0.60

โปรแกรมสำเร็จรูปจะทำการประมาณค่าของความน่าจะเป็นที่ case แต่ละ case ในกลุ่มตัวอย่างจะเข้าข่ายมี ses อยู่ในระดับต่ำ (EST_1) กลาง (EST2_1) สูง (EST3_1) ค่า ses ที่ได้จากการพยากรณ์ (PRE_1) ค่าความน่าจะเป็นที่พยากรณ์ได้ (PCP_1) และความน่าจะเป็นของ ses จริง ตัวอย่างเช่น case ที่ 1 จากกลุ่มตัวอย่างจะมี ses=1(ต่ำ) ในขณะที่ OLOGIT จะพยากรณ์ค่าความ

น่าจะเป็นที่ case ที่ 1 จะมี ses ต่ำ(1) กลาง(2) สูง(3) เท่ากับ 0.16 , 0.50 และ 0.34 ตามลำดับ เนื่องด้วยค่าความน่าจะเป็นของ case ที่ 1 ที่จะมี ses อยู่ในระดับกลาง(2) สูงที่สุด ดังนั้นOLOGIT จึงพยากรณ์ว่า case ที่ 1 มี ses อยู่ในระดับกลาง (2) แต่แท้จริงแล้ว case ที่ 1 มี ses อยู่ในระดับต่ำ (1)

จากข้อมูลของ ses จริง และ ses ที่ได้จากการพยากรณ์ นักสถิติสามารถนำมาทำตารางไขว้ (cross tabulation) ปรากฏผลดังนี้

ses * Predicted Response Category Crosstabulation

Count		Predicted Response Category			Total
		low	middle	high	
ses	low	11	35	1	47
	middle	8	74	13	95
	high	1	39	18	58
Total		20	148	32	200

ความถูกต้องของการพยากรณ์ =
 $(11+74+18) \times 100 / (200)$
 =51.5

ซึ่งเมื่อกำหนดความถูกต้องของการพยากรณ์ พบว่ามีความถูกต้องเพียงร้อยละ 51.5 เท่านั้น สอดคล้องกับดัชนี

วัด goodness-of-fit อื่น ๆ เช่น pseudo-R-square ที่มีค่าอยู่ในระดับต่ำ

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 หลีกเลี่ยงความผิดพลาดเป็นร้อยเป็นพันล้านบาทเกิดการลงทุนจริง(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 ข้อครอบคลุมเนื้อหาในระดับอุดมศึกษาและTOEFL) มาพร้อมกับไฟล์เสียง/ไฟล์ข้อมูล/ฯลฯ อีกมาก(กำหนด DEBUT 1 เมษายน 2569)