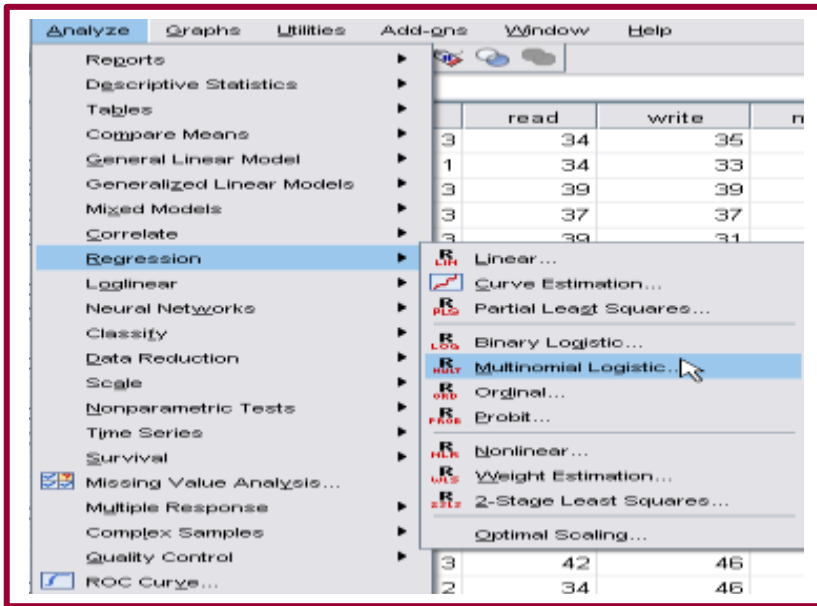


Multinomial Logistic Regression Workshop #1-Drop-down menu

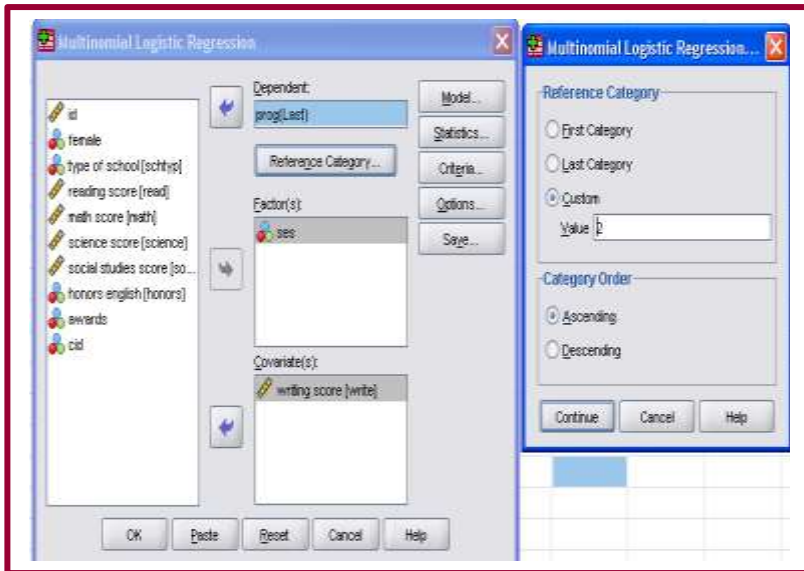
ตารางด้านล่างแสดงข้อมูลการสมัครเข้าเรียนในหลักสูตรต่าง ๆ ของนักศึกษา ซึ่งผู้วิเคราะห์สนใจที่จะใช้ดัชนีแสดงความเป็นอยู่ทางสังคม (social economic status) ซึ่งเป็นตัวแปรนามบัญญัติสามระดับและคะแนนการสอบข้อเขียนในการพยากรณ์หลักสูตรที่นักศึกษาเลือกเข้าศึกษา

	id	female	ses	schtyp	prog	read	write	math	science	socst	honors	awards	cid
1	45	1	1	1	3	34	35	41	29	26	0	0	1
2	108	0	2	1	1	34	33	41	36	36	0	0	1
3	15	0	3	1	3	39	39	44	26	42	0	0	1
4	67	0	1	1	3	37	37	42	33	32	0	0	1
5	153	0	2	1	3	39	31	40	39	51	0	0	1
6	51	1	3	1	1	42	36	42	31	39	0	0	1
7	164	0	2	1	3	31	36	46	39	46	0	0	1
8	133	0	2	1	3	50	31	40	34	31	0	0	1
9	2	1	2	1	3	39	41	33	42	41	0	0	1
10	53	0	2	1	3	34	37	46	39	31	0	0	1
11	1	1	1	1	3	34	44	40	39	41	0	0	1
12	128	0	3	1	2	39	33	38	47	41	0	0	2
13	16	0	1	1	3	47	31	44	36	38	0	0	2
14	106	1	2	1	3	36	44	37	42	41	0	0	2
15	89	1	1	1	3	36	36	40	61	33	0	0	2
16	134	0	1	1	1	44	44	39	34	46	0	0	2
17	19	1	1	1	1	28	46	43	44	51	0	0	2
18	146	1	2	1	3	42	46	38	36	46	0	0	2
19	11	0	2	1	2	34	46	46	39	36	0	0	2
20	117	0	3	1	3	34	49	39	42	66	0	0	2
21	109	1	2	1	1	42	39	42	42	41	0	0	2
22	12	0	2	1	3	37	44	46	39	46	0	0	3
23	37	1	1	1	3	41	47	40	39	61	0	0	3
24	69	1	1	1	3	44	44	40	40	31	0	0	3
25	43	1	1	1	2	47	37	43	42	46	0	0	3
26	196	0	3	2	2	44	38	49	39	46	0	0	3
27	36	1	1	1	1	44	49	44	35	51	0	0	3
28	165	0	2	1	1	44	44	46	39	51	0	0	3
29	6	1	1	1	2	47	41	46	40	41	0	0	3
30	4	1	1	1	2	44	50	41	39	51	0	1	3
31	25	1	2	1	1	47	44	42	42	36	0	0	4
32	107	0	1	1	3	47	39	47	42	28	0	0	4
33	5	0	1	1	2	47	40	43	46	31	0	0	4
34	47	1	1	1	2	47	46	49	33	41	0	0	4
35	140	0	2	1	3	44	41	40	50	26	0	0	4
36	22	0	2	1	3	42	39	39	56	46	0	0	4
37	18	0	2	1	3	50	33	49	44	36	0	0	4
38	30	1	3	1	2	41	59	42	34	51	0	2	4
39	40	0	1	1	1	42	41	43	50	41	0	0	4

•คลิก Analyze>Regression>Multinomial logistic

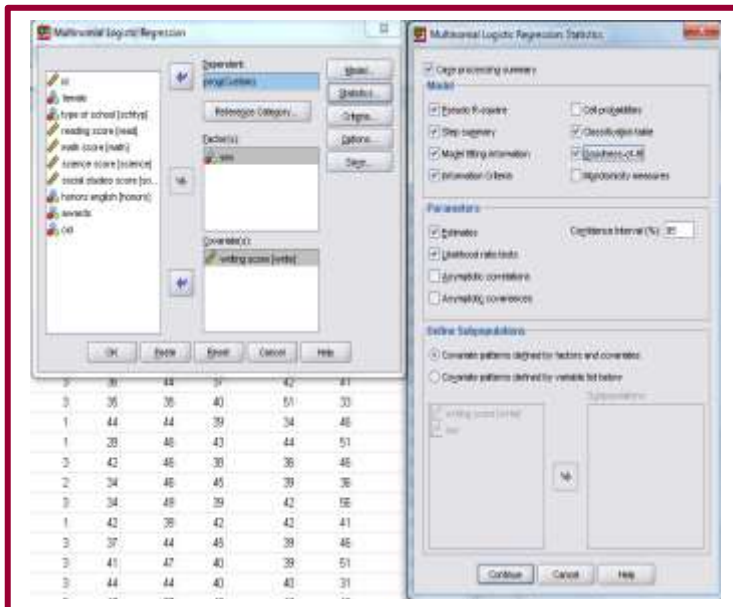


จะปรากฏ menu ดังต่อไปนี้



•คลิกเลือก ses (social economic status-ดัชนีแสดงฐานะทางเศรษฐกิจสังคม) ให้ไปอยู่ในกล่อง Factor(s) write ไปอยู่ในกล่อง Covariate(s) คลิกเลือก prog (หลักสูตรที่เลือกเรียน) ไปอยู่ในกล่อง Dependent คลิกเลือกให้ prog=2 เป็นหลักสูตรอ้างอิงเมื่อต้องมีการนำหลักสูตรอื่นมา

•คลิกที่ปุ่ม Statistics จะปรากฏ menu



●คลิกเลือก Case processing summary
 ในส่วน Model คลิกเลือก Pseudo R-square, Step summary , Model fitting information, Information Criteria ,Classification table, Goodness-of-fit ในส่วน Parameters คลิกเลือก Estimates และ Likelihood ratio tests

ต่อไปนี้เป็น computer output ที่ได้

Case Processing Summary

		N	Marginal Percentage
type of program	general	45	22.5%
	academic	105	52.5%
	vocation	50	25.0%
ses	low	47	23.5%
	middle	95	47.5%
	high	58	29.0%
Valid		200	100.0%
Missing		0	
Total		200	
Subpopulation		64 ^a	

a. The dependent variable has only one value observed in 32 (50.0%) subpopulations.

ขนาดของกลุ่มตัวอย่างเท่ากับ 200 ประกอบด้วยผู้ที่เลือกเรียนหลักสูตรสามัญ 45 ราย วิชาการ 105 ราย อาชีวะ 50 ราย หากแยกตามสถานะทางเศรษฐกิจสังคม พบว่า อยู่ในระดับล่าง 47 ราย ระดับกลาง 95 ราย และระดับบน 58 ราย

ตารางต่อไปเป็นตาราง model fitting information

Model Fitting Information

Model	Model Fitting Criteria			Likelihood Ratio Tests		
	AIC	BIC	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	258.986	265.583	254.986			
Final	222.756	249.143	206.756	48.230	6	.000

ตาราง Model Fitting Information แสดงดัชนีหลายอย่างที่จะช่วยให้นักวิเคราะห์สามารถใช้ประเมินแบบจำลองที่มีแต่ intercept (หรือบางทีเรียกว่า null model) กับแบบจำลองที่รวมตัวแปรอิสระเข้ากับ intercept (ที่เรียกกันว่า full model หรือ final model) ค่าของ Akaike Information Criterion (AIC) และ Bayesian Information Criterion (BIC) เป็นสถิติที่ใช้ในการวัดความถูกต้องของแบบจำลอง (model fit) ยิ่งค่าสถิติทั้งสองมีค่าต่ำเท่าใดก็ยิ่งแสดงความ fit ของ model มากขึ้นเท่านั้น (จะเป็นประโยชน์ก็ต่อเมื่อมีค่าของ model อื่นมาเปรียบเทียบ) หากแบบจำลองใช้ได้ผล -2 Log Likelihood (-2LL) สำหรับ full model ที่ได้จะมีค่าต่ำกว่า null model ยิ่งมีค่าต่ำกว่ามากเท่าใดก็ยิ่งแสดงให้เห็นว่าแบบจำลองที่ใช้ยิ่งมีความถูกต้องเที่ยงตรง (model fit) มากขึ้นเท่านั้น -2LL เป็น likelihood ratio และแสดงความผันผวนของตัวแปรตามที่ไม่สามารถอธิบายได้ด้วยตัวแปรอิสระ ดังนั้นยังมีค่าต่ำยิ่งดี ส่วน Likelihood Ratio chi-square test ก็เป็นอีกทางเลือกหนึ่งของการทดสอบเพื่อดู goodness-of-fit แต่มีแนวโน้มที่จะมีค่าสูงโดยเฉพาะหากกลุ่มตัวอย่างมีขนาดใหญ่ ในที่นี้ค่าไคร้สแควร์ที่คำนวณได้เท่ากับ 48.230 และ p-value ต่ำกว่า 0.05 ซึ่งบ่งบอกว่าการนำตัวแปรอิสระเข้ามาช่วยอธิบายความผันผวนในตัวแปรตามได้อย่างมีนัยสำคัญ

ตารางถัดไปแสดงตาราง goodness-of-fit

	Chi-Square	df	Sig.
Pearson	119.766	120	.489
Deviance	129.875	120	.254

ตาราง goodness-of-fit เป็นอีกตารางหนึ่งที่แสดงหลักฐานให้เห็นว่าค่าความถี่ที่ได้จากการสังเกต (observed frequency) มีค่าใกล้เคียงกับค่าความถี่ที่ได้จากการประมาณการ (expected frequency) และแบบจำลองมีความถูกต้อง ในที่นี้ค่า Pearson Chi-square และ Deviance Chi-square จะต้องไม่มีนัยสำคัญทางสถิติ จึงจะสามารถตีความได้ว่าแบบจำลองมีความถูกต้อง (model fit) อนึ่งค่าสถิติทั้งสองมีแนวโน้มที่จะมีค่าสูง (มีแนวโน้มที่จะมีนัยสำคัญทางสถิติ) หากกลุ่มตัวอย่างมีขนาดใหญ่

Cox and Snell	.214
Nagelkerke	.246
McFadden	.118

ในที่นี้ Pseudo R-Square มีค่าอยู่ในช่วง 0.1-0.2 แสดงว่าแบบจำลองที่มีการรวมตัวแปรอิสระเข้าไว้ด้วย ช่วยอธิบายตัวแปรตามได้บ้างแม้จะไม่สูงนัก

Effect	Model Fitting Criteria			Likelihood Ratio Tests		
	AIC of Reduced Model	BIC of Reduced Model	-2 Log Likelihood of Reduced Model	Chi-Square	df	Sig.
Intercept	222.756	249.143	2.068E2 ^a	.000	0	.
write	250.203	269.993	238.203	31.447	2	.000
ses	225.815	239.008	217.815	11.058	4	.026

The chi-square statistic is the difference in -2 log-likelihoods between the final model and a reduced model. The reduced model is formed by omitting an effect from the final model. The null hypothesis is that all parameters of that effect are 0.

a. This reduced model is equivalent to the final model because omitting the effect does not increase the degrees of freedom.

ตาราง Likelihood Ratio Tests แสดงค่าสถิติต่าง ๆ เหมือนเช่นในตาราง Model Fitting Information แต่ในที่นี้เป็น การแสดงให้เห็นแต่ละองค์ประกอบไม่ว่าจะเป็นตัวแปรอิสระที่เป็นตัวแปรนามบัญญัติ (ses) และตัวแปรอิสระที่มีค่าต่อเนื่อง (write) ในการอธิบาย goodness-of-fit ของแบบจำลองที่ใช้ ในที่นี้ค่า p-value ของตัวแปรทั้งสองประเภทมีค่าต่ำกว่า 0.05 ดังนั้น นักสถิติจำเป็นต้องอาศัยตัวแปรทั้งสองประเภทในการพยากรณ์ค่าของตัวแปรตาม และไม่สามารถตัดตัวแปรตัวใดตัวหนึ่งออกไปได้

type of program ^a	B	Std. Error	Wald	df	Sig.	Exp(B)	95% Confidence Interval for Exp(B)	
							Lower Bound	Upper Bound
general	Intercept	1.689	1.227	1.896	1	.169		
	write	-.058	.021	7.320	1	.007	.944	.905 .984
	[ses=1]	1.163	.514	5.114	1	.024	3.199	1.168 8.764
	[ses=2]	.630	.465	1.833	1	.176	1.877	.754 4.669
	[ses=3]	0 ^b			0			
vocation	Intercept	4.236	1.205	12.361	1	.000		
	write	-.114	.022	26.139	1	.000	.893	.855 .932
	[ses=1]	.983	.596	2.722	1	.099	2.672	.831 8.585
	[ses=2]	1.274	.511	6.214	1	.013	3.575	1.313 9.736
	[ses=3]	0 ^b			0			

a. The reference category is: academic.

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

ตาราง Parameter Estimates แสดงให้เห็นว่า

- หากคะแนน write เพิ่มขึ้นหนึ่งหน่วยมีผลทำให้ log -odds ของการเลือก general program แทนที่จะเป็น academic program ลดลง 0.058
- หากคะแนน write เพิ่มขึ้นหนึ่งหน่วยมีผลทำให้ log-odds ของการเลือก vocation program เมื่อเทียบกับ academic program ลดลง 0.114
- หากสถานะทางเศรษฐกิจสังคมลดลงจากระดับสูงสุดคือ 3 ลงไประดับล่างสุด มีผลทำให้ log-odds ของการเลือก general program เมื่อเทียบกับ academic program เพิ่มขึ้น 1.163
- หากคะแนน write เพิ่มขึ้นหนึ่งหน่วยมีผลทำให้ odd ของการเลือก general program แทนที่จะเป็น academic program ลดลงร้อยละ 5.6 (.944-1)
- หากสถานะทางเศรษฐกิจสังคมลดลงจาก 3 เป็น 1 มีผลทำให้ odd ของการเลือก general program แทนที่จะเป็น academic program เพิ่มขึ้นร้อยละ 219.9 (3.199-1)

Classification

Observed	Predicted			Percent Correct
	general	academic	vocation	
general	7	27	11	15.6%
academic	4	92	9	87.6%
vocation	4	23	23	46.0%
Overall Percentage	7.5%	71.0%	21.5%	61.0%

Classification table แสดงให้เห็นความแม่นยำในการพยากรณ์โดยเฉลี่ยเท่ากับ 61.0 %

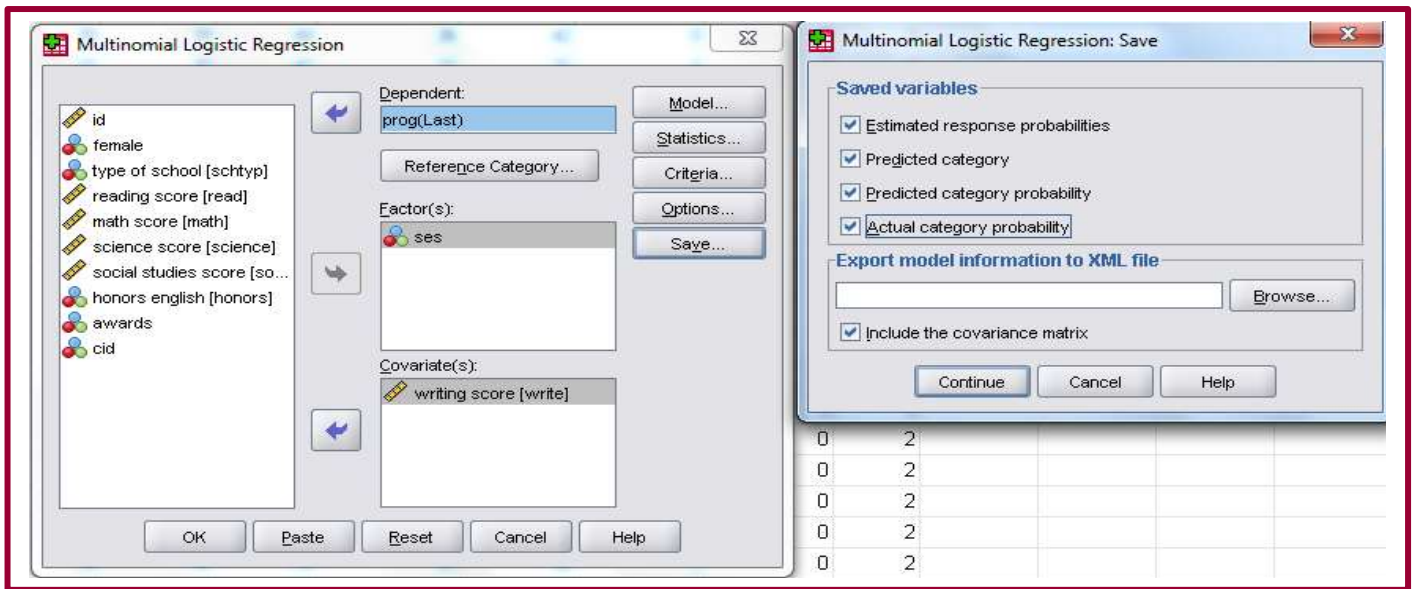
หากใช้ข้อมูลของหลักสูตร (program) ที่นักศึกษาเลือก เราสามารถคำนวณสัดส่วนที่เราจะเดา (สัดส่วนที่เกิดจากโอกาส) ได้ถูกต้องดังนี้

$$\begin{aligned} & \text{สัดส่วนที่สามารถเดาได้ถูกต้องโดยไม่ใช้แบบจำลองใด ๆ} = \text{ผลบวกของสัดส่วนยกกำลังสอง} \\ & = (.225)^2 + (.525)^2 + (.25)^2 \\ & = .3887 \end{aligned}$$

หากให้สัดส่วนที่เดาได้ถูกต้องเพิ่มขึ้นร้อยละ 25% สัดส่วนที่จะเดาได้ถูกต้องคือ .485875(1.25X.3887) หรือร้อยละ 48.58

ดังนั้นจากการใช้ MLR เราสามารถพยากรณ์ค่าของตัวแปรตามได้ถูกต้องถึงร้อยละ 61.0 ซึ่งสูงกว่าสัดส่วนความถูกต้องที่ได้จากการเดาหรือจากโอกาสมากพอสมควร

หากในเมนูหลักมีการกดปุ่ม save และเลือกตัวแปรที่ต้องการ save ค่าไว้ดังภาพต่อไปนี้



โดย Estimated response probabilities คือความน่าจะเป็นของตัวแปรตาม ที่มีได้หลายค่าตามจำนวนประเภท หรือกลุ่มที่มีการกำหนดขึ้นในตัวแปรตาม

Predicted category คือ ประเภทหรือกลุ่มของตัวแปรตาม พิจารณาจาก estimated response probabilities

Predicted category probability คือค่าความน่าจะเป็นที่สอดคล้องกับ predicted category

Actual category probability คือค่าความน่าจะเป็นของประเภทหรือกลุ่มของตัวแปรตามที่มีการเกิดขึ้นจริง

โปรแกรมสำเร็จรูปจะแสดงผลเพิ่มขึ้นอีก 6 คอลัมน์ แสดงได้ดังภาพต่อไปนี้

ตัวแปรตาม

ความน่าจะเป็นที่ตัวแปรตามจะมีค่า 1/2/3

ค่าของตัวแปรตามที่ได้จากการพยากรณ์ พร้อมค่าความน่าจะเป็น

ความน่าจะเป็นของค่าตัวแปรตามที่เกิดขึ้นจริง

	prog	read	write	math	science	socst	honors	awards	cid	EST1_1	EST2_1	EST3_1	PRE_1	PCP_1	ACP_1
1	3	34	35	41	29	26	0	0	1	0.34	0.15	0.51	3	0.51	0.51
2	1	34	33	41	36	36	0	0	1	0.18	0.12	0.70	3	0.70	0.18
3	3	39	39	44	26	42	0	0	1	0.24	0.42	0.34	2	0.42	0.34
4	3	37	37	42	33	32	0	0	1	0.36	0.17	0.48	3	0.48	0.48
5	3	39	31	40	39	51	0	0	1	0.17	0.10	0.73	3	0.73	0.73
6	1	42	36	42	31	39	0	0	1	0.24	0.36	0.41	3	0.41	0.24
7	3	31	36	46	39	46	0	0	1	0.20	0.16	0.65	3	0.65	0.65
8	3	50	31	40	34	31	0	0	1	0.17	0.10	0.73	3	0.73	0.73
9	3	39	41	33	42	41	0	0	1	0.22	0.23	0.55	3	0.55	0.55
10	3	34	37	46	39	31	0	0	1	0.20	0.17	0.63	3	0.63	0.63
11	3	34	44	40	39	41	0	0	1	0.38	0.28	0.35	1	0.38	0.35
12	2	39	33	38	47	41	0	0	2	0.23	0.29	0.47	3	0.47	0.29
13	3	47	31	44	36	36	0	0	2	0.31	0.11	0.58	3	0.58	0.58
14	3	36	44	37	42	41	0	0	2	0.23	0.29	0.48	3	0.48	0.48
15	3	35	35	40	51	33	0	0	2	0.34	0.15	0.51	3	0.51	0.51
16	1	44	44	39	34	46	0	0	2	0.38	0.28	0.35	1	0.38	0.38
17	1	28	46	43	44	51	0	0	2	0.38	0.31	0.31	1	0.38	0.38
18	3	42	46	38	36	46	0	0	2	0.23	0.33	0.44	3	0.44	0.44
19	2	34	46	45	39	36	0	0	2	0.23	0.33	0.44	3	0.44	0.33
20	3	34	49	39	42	56	0	0	2	0.20	0.63	0.17	2	0.63	0.17
21	1	42	39	42	42	41	0	0	2	0.21	0.20	0.59	3	0.59	0.21
22	3	37	44	45	39	46	0	0	3	0.23	0.29	0.48	3	0.48	0.48
23	3	41	47	40	39	51	0	0	3	0.38	0.33	0.29	1	0.38	0.29
24	3	44	44	40	40	31	0	0	3	0.38	0.28	0.35	1	0.38	0.35

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