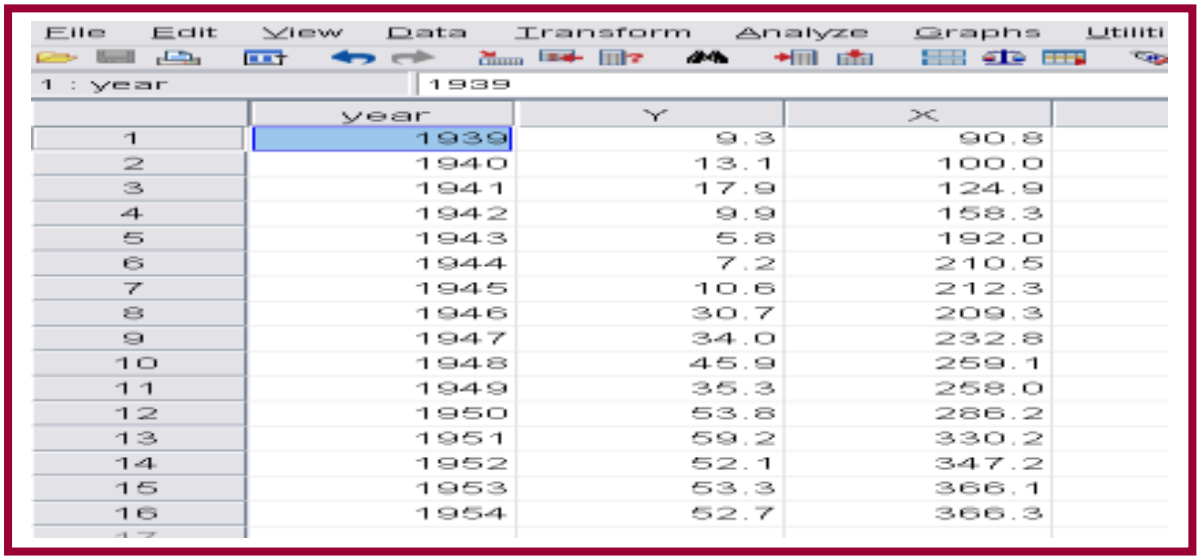
Multiple Regression Analysis Workshop #2 -Use of dummy variables

ตารางต่อไปนี้แสดงข้อมูลของ gross private domestic investment (แทนด้วยตัวแปร Y) และ gross national product (แทนด้วยตัวแปร X) ในช่วงระหว่างปี 1939 ถึงปี 1954 ซึ่งเป็นตัวอย่างสาธิตการใช้ตัวแปรหุ่น (dummy variable) ในการวิเคราะห์การถดถอยพหุคูณ



File Edit View Data Transform Analyze Graphs Utiliti

1 : year 1939

	year	Y	X
1	1939	9.3	90.8
2	1940	13.1	100.0
3	1941	17.9	124.9
4	1942	9.9	158.3
5	1943	5.8	192.0
6	1944	7.2	210.5
7	1945	10.6	212.3
8	1946	30.7	209.3
9	1947	34.0	232.8
10	1948	45.9	259.1
11	1949	35.3	258.0
12	1950	53.8	286.2
13	1951	59.2	330.2
14	1952	52.1	347.2
15	1953	53.3	366.1
16	1954	52.7	366.3

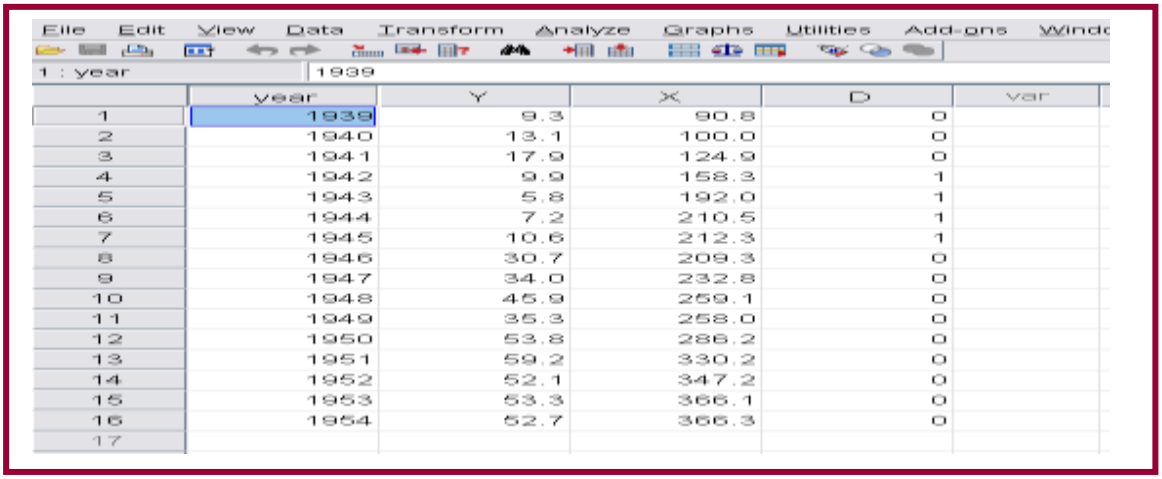
ในการหาสมการถดถอยพหุคูณโดยใช้ Y เป็นตัวแปรตาม และ X เป็นตัวแปรต้นให้มีประสิทธิภาพในแง่ของการพยากรณ์ข้อมูลที่เป็น chronological order นั้น จำเป็นต้องคำนึงถึงสภาพแวดล้อมสังคม เศรษฐกิจ การเมือง และเทคโนโลยีที่ค่อนข้างเหมือนเดิม ไม่มีการเปลี่ยนแปลง แต่ในตัวอย่างข้างต้นนี้มีบางปีที่อยู่ในช่วงการมีศึกสงคราม (1942-1945) เป็นช่วงที่มีการสะดุด หรือdisruption เกิดขึ้น ดังนั้นระดับกิจกรรมต่าง ๆ ที่ดำเนินอยู่ในช่วงศึกสงคราม อาจมีความแตกต่างอย่างมากจากกิจกรรมในยามสงบ เมื่อเป็นเช่นนี้ เราอาจใช้ตัวแปรหุ่น (dummy variable) เข้ามาแยกแยะให้ข้อมูลช่วงที่มีความสงบแตกต่างจากช่วงที่มีศึกสงคราม

ตารางด้านล่างมีการเพิ่มตัวแปรหุ่น (dummy variable) เข้ามา ใช้สัญลักษณ์ D โดย

D=0 ในปีที่เป็นช่วงสงบ

D=1 ในปีที่มีศึกสงคราม

ผลจากการเพิ่มตัวแปรหุ่นเข้ามาทำให้เราได้ตารางข้อมูลใหม่เป็นดังนี้



	year	Y	X	D	var
1	1939	9.3	90.8	0	
2	1940	13.1	100.0	0	
3	1941	17.9	124.9	0	
4	1942	9.9	158.3	1	
5	1943	5.8	192.0	1	
6	1944	7.2	210.5	1	
7	1945	10.6	212.3	1	
8	1946	30.7	209.3	0	
9	1947	34.0	232.8	0	
10	1948	45.9	259.1	0	
11	1949	35.3	258.0	0	
12	1950	53.8	286.2	0	
13	1951	59.2	330.2	0	
14	1952	52.1	347.2	0	
15	1953	53.3	366.1	0	
16	1954	52.7	366.3	0	
17					

สมการถดถอยที่เราต้องการจะอยู่ในรูปต่อไปนี้

$Y = a_O + a_1 X + a_2 D$

โดยที่ Y=gross private domestic investment

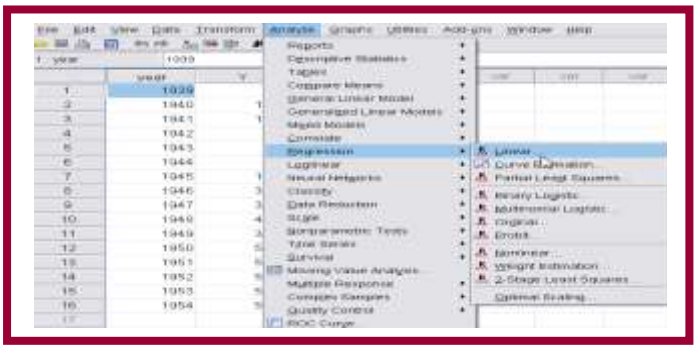
X=gross national product

D=dummy variable

a_O= y-intercept

a_i = regression coefficients (i=1,2)

◉ที่ menu หลัก คลิ๊ก Analyze>Regression>Linear แสดงได้ดังภาพที่ 1 ต่อไปนี้



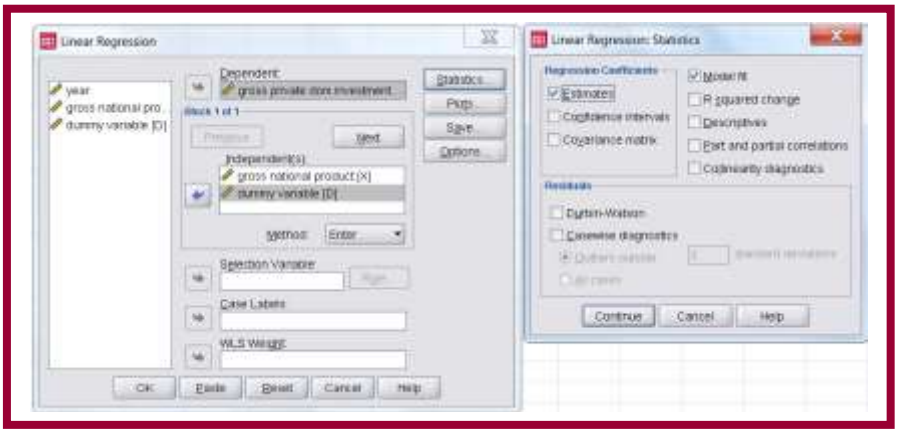
ภาพที่ 1

Submenu จะปรากฏขึ้นดังในภาพที่ 2



ภาพที่ 2

◉ คลิ๊กเลือก Y ให้ไปอยู่ในช่อง Dependent คลิ๊กเลือก X และ D ให้ไปอยู่ในช่อง Independent(s) จากนั้นคลิ๊กปุ่ม Statistics จะปรากฏ submenu ด้านขวาของภาพที่ 3 ให้คลิ๊ก Estimates ภายใต้หัวข้อ Regression Coefficients และคลิ๊ก Model fit



ภาพที่ 3

◉คลิ๊ก Continue ท้าย submenu Statistics จากนั้นคลิ๊ก Ok ท้าย menu Linear Regression ภาพที่ 4 ขวามือแสดง computer output ที่ได้จากการ run program

ภาพที่ 4

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.972[a]	.944	.935	5.0876

a. Predictors: (Constant), dummy variable, gross national product

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	5665.639	2	2832.820	109.443	.000[a]
	Residual	336.491	13	25.884		
	Total	6002.130	15			

a. Predictors: (Constant), dummy variable, gross national product

b. Dependent Variable: gross private dom.investment

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	-2.574	4.046		-.636	.536
	gross national product	.164	.015	.736	10.790	.000
	dummy variable	-20.811	3.052	-.465	-6.820	.000

a. Dependent Variable: gross private dom.investment

เราสามารถประมวลผล computer output ได้ดังนี้

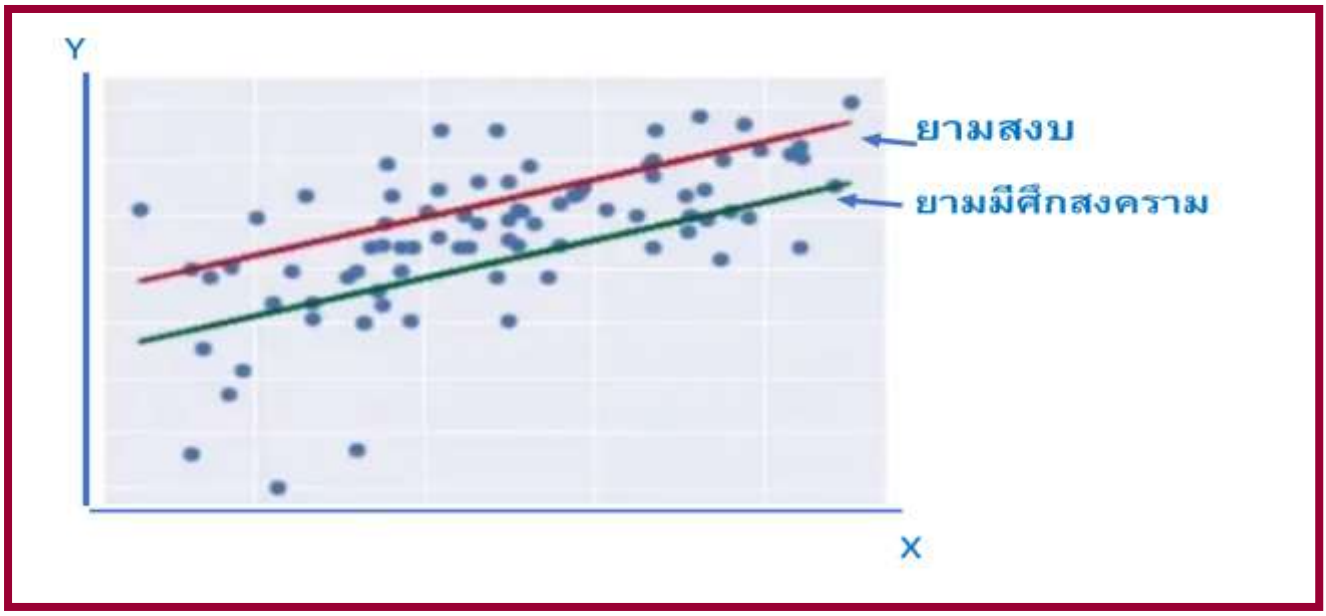
1) Coefficient of multiple determination , R^2 มีค่า 0.944 บ่งบอกว่า gross national product และ dummy Variable สามารถอธิบายความผันผวนใน gross private domestic investment ได้ถึงร้อยละ 94.4
2) สมการถดถอยที่ต้องการคือ y=-2.574 +.164X -20.811D

   โดยที่ Y= gross private domestic investment(การลงทุนภาคเอกชนภายในประเทศมวลรวม)

   X= gross national product(ผลิตภัณฑ์มวลรวมประชาชาติ)

   D=ตัวแปรหุ่น(=0 ในยามปกติ ,=1 ในยามสงคราม)

ผลจากการมีตัวแปรหุ่น เป็นผลให้มีการเปลี่ยนแปลงในเส้น regression โดยมีการเพิ่มหรือลดใน y-intercept แต่ไม่มีผลกระทบต่อ slope ของเส้น regression แสดงได้ดังภาพล่างต่อไปนี้



หากเราใช้ข้อมูลข้างต้นโดยงดเว้นไม่ใช้ตัวแปรหุ่น (dummy variable) เราจะได้ผลลัพท์ดังนี้

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.862[a]	.743	.725	10.4892

a. Predictors: (Constant), gross national product

b. Dependent Variable: gross private dom.investment

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	4461.801	1	4461.801	40.553	.000[a]
	Residual	1540.329	14	110.023		
	Total	6002.130	15			

a. Predictors: (Constant), gross national product

b. Dependent Variable: gross private dom.investment

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	-14.362	7.543		-1.904	.079
	gross national product	.192	.030	.862	6.368	.000

a. Dependent Variable: gross private dom.investment

จะเห็นได้ว่าค่า R^2 กรณีนี้มีค่าต่ำกว่า R^2 กรณีที่มีการใช้ dummy variable นั่นหมายความว่า การเพิ่มdummy variable เข้ามาในแบบจำลองมีผลทำให้ความแม่นยำของการพยากรณ์สูงขึ้น

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 หลีกเลี่ยงความผิดพลาดเป็นร้อยเป็นพันล้านหากเกิดการลงทุนจริง(กำหนด DEBUT 1 เมษายน 2569)

* ผู้แต่งหนังสือ"การวิเคราะห์ความเป็นไปได้ทางการเงินและการจัดวงเงินเครดิตของโครงการลงทุน"ประกอบด้วยตัวอย่างของธุรกิจจริงที่ไม่เปิดเผยชื่อนับ 100 บริษัท ครอบคลุมอุตสาหกรรม 24 อุตสาหกรรม

*Co-developer ซอฟแวร์ 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)