

บทที่ 5

การวิเคราะห์ห้กดถอย เชิงเส้นและสหสัมพันธ์

ผศ. รินทร์หทัย กิตติธนาอรุณ





สำหรับเนื้อหาในบทนี้ ประกอบด้วย

- ❖ สมการถดถอยเชิงเส้นอย่างง่าย
- ❖ การหาค่าสัมประสิทธิ์สหสัมพันธ์
- ❖ การใช้โปรแกรมสำเร็จรูปในการหาสมการถดถอยเชิงเส้นอย่างง่าย
สมการถดถอยพหุคูณ และ การหาค่าสัมประสิทธิ์สหสัมพันธ์



การวิเคราะห์การถดถอยเป็นวิธีการทางสถิติ ที่ศึกษาถึงความสัมพันธ์ระหว่าง ตัวแปรตัวหนึ่ง ซึ่งเรียกว่า “ตัวแปรตาม (Dependent variable) แทนด้วย Y ” และตัวแปรอื่น ๆ ซึ่งเรียกว่า “ตัวแปรอิสระ (Independent variable) แทนด้วย X ”ว่ามีความสัมพันธ์กันอย่างไร ซึ่งความสัมพันธ์ดังกล่าวนี้จะอยู่ในรูปแบบคณิตศาสตร์โดยอาจจะเป็นเชิงเส้นตรงก็ได้หรือไม่เป็นก็ได้ มีประโยชน์เพื่อใช้ในการทำนาย (Prediction) นั่นคือ ใช้ตัวแปรอิสระ X เป็นเกณฑ์ในการประมาณตัวแปรตาม Y โดยสำหรับในบทนี้จะกล่าวถึง การวิเคราะห์การถดถอยที่เป็นเชิงเส้นอย่างง่ายเท่านั้น



❖ สมการถดถอยเชิงเส้นอย่างง่าย

ตามรูปแบบสมการถดถอยเชิงเส้นอย่างง่าย $Y = a + bX$ ตัวแปร Y จะมีค่ามากหรือน้อยขึ้นอยู่กับตัวแปร X

โดยเรียก ตัวแปร Y ว่าตัวแปรตาม

ตัวแปร X ว่าตัวแปรอิสระ

a เป็นระยะตัดแกน Y ของเส้นแนวโน้ม (y-intercept)

b เป็นความชัน (slope) ของเส้นแนวโน้ม ซึ่ง หมายถึงอัตราการเปลี่ยนแปลงของ Y เมื่อ X เปลี่ยนแปลงไป 1 หน่วย

สำหรับการหาสมการเส้นตรงหรือหาค่า a และ b วิธีการที่นิยมใช้อย่างแพร่หลายก็คือ วิธีกำลังสองน้อยที่สุด (Least square method) ดังนี้

$$b = \frac{\sum XY - n\bar{X}\bar{Y}}{\sum X^2 - n\bar{X}^2}, \quad a = \bar{Y} - b\bar{X}$$



ตัวอย่างที่ 1 ตารางแสดงข้อมูลเปอร์เซ็นต์เกษตรกรและค่าแตกต่างของรายได้

เปอร์เซ็นต์เกษตรกร (X)	ค่าแตกต่างของ รายได้ (Y)	X^2	XY
2.13	809	4.5369	1723.17
2.52	763	6.3504	1922.76
11.86	612	140.6596	7258.32
2.55	492	6.5025	1254.6
2.87	679	8.2369	1948.73
4.23	635	17.8929	2686.05



เปอร์เซ็นต์เกษตรกร (X)	ค่าแตกต่างของรายได้ (Y)	X^2	XY
4.62	859	21.3444	3968.58
5.19	228	26.9361	1183.32
6.43	897	41.3449	5767.71
6.70	867	44.89	5808.9
1.53	513	2.3409	784.89
1.87	335	3.4969	626.45
10.38	868	107.7444	9009.84
62.88	8,557	432.2768	43943.32



จากสูตร

$$b = \frac{\sum XY - n\bar{X}\bar{Y}}{\sum X^2 - n\bar{X}^2}$$

$$\bar{X} = \frac{\sum X}{n} = \frac{62.88}{13} = 4.837$$

$$\bar{Y} = \frac{\sum Y}{n} = \frac{8,557}{13} = 658.231$$

$$\therefore b = \frac{43943.32 - (13)(4.837)(658.231)}{432.2768 - (13)(4.837)^2}$$

$$= \frac{43943.32 - 41390.2235}{432.2768 - 304.1553}$$



$$= \frac{2553.0965}{128.1215} = 19.927$$

$$a = \bar{Y} - b\bar{X}$$

$$= 658.231 - (19.927)(4.837)$$

$$= 658.231 - 96.387$$

$$= 561.844$$

จะได้สมการถดถอยเชิงเส้นตรงดังนี้ $Y = 561.844 + 19.927X$



ตัวอย่างที่ 2 จากข้อมูลแสดงรายได้ต่อเดือนและราคาห้องพักของนักท่องเที่ยวที่ได้จากการสำรวจ
ในจังหวัดบุรีรัมย์ เป็นดังนี้

นักท่องเที่ยว	รายได้/เดือน	ราคาห้องพัก
1	3,000	225
2	3,560	200
3	4,120	459
4	4,300	445
5	4,870	439
6	5,230	577
7	5,400	722
8	5,750	903
9	5,980	1,350
10	6,350	1,360

จงหาสมการถดถอย
เส้นตรงอย่างง่าย



❖ การหาค่าสัมประสิทธิ์สหสัมพันธ์

ในการศึกษาเกี่ยวกับสมการเส้นถดถอยนั้น มีวัตถุประสงค์เพื่อที่จะคาดคะเนค่าของตัวแปรสุ่มตัวหนึ่งในขณะที่ทราบค่าของตัวแปรอีกตัวหนึ่ง การคาดคะเนดังกล่าวจะแม่นยำมากน้อยเพียงใดนั้นขึ้นอยู่กับปริมาณความสัมพันธ์ระหว่างตัวแปรคู่ (X, Y) ซึ่งจะเรียกว่าสัมประสิทธิ์สหสัมพันธ์

ข้อมูลที่จะศึกษาสัมประสิทธิ์สหสัมพันธ์ จะต้องเป็นตัวแปรสุ่ม 2 ตัว ซึ่งเป็นตัวแปรคู่ที่ได้จากหน่วยทดลองเดียวกัน ค่าสัมประสิทธิ์สหสัมพันธ์ ใช้สัญลักษณ์ r ค่าสัมประสิทธิ์สหสัมพันธ์มีค่าระหว่าง -1.00 ถึง $+1.00$ ซึ่งโดยทั่วไปมีความหมาย ดังนี้



ค่าสัมประสิทธิ์สหสัมพันธ์ (r)

แสดงว่า

+1.00

มีความสัมพันธ์กันอย่างสมบูรณ์ทางบวก

+0.85

มีความสัมพันธ์กันค่อนข้างสูงทางบวก

+0.50

มีความสัมพันธ์กันปานกลางทางบวก

+0.20

มีความสัมพันธ์กันค่อนข้างต่ำทางบวก

0.00

ไม่มีความสัมพันธ์กันเลย

-0.20

มีความสัมพันธ์กันค่อนข้างต่ำทางลบ

-0.50

มีความสัมพันธ์กันปานกลางทางลบ

-0.85

มีความสัมพันธ์กันค่อนข้างสูงทางลบ

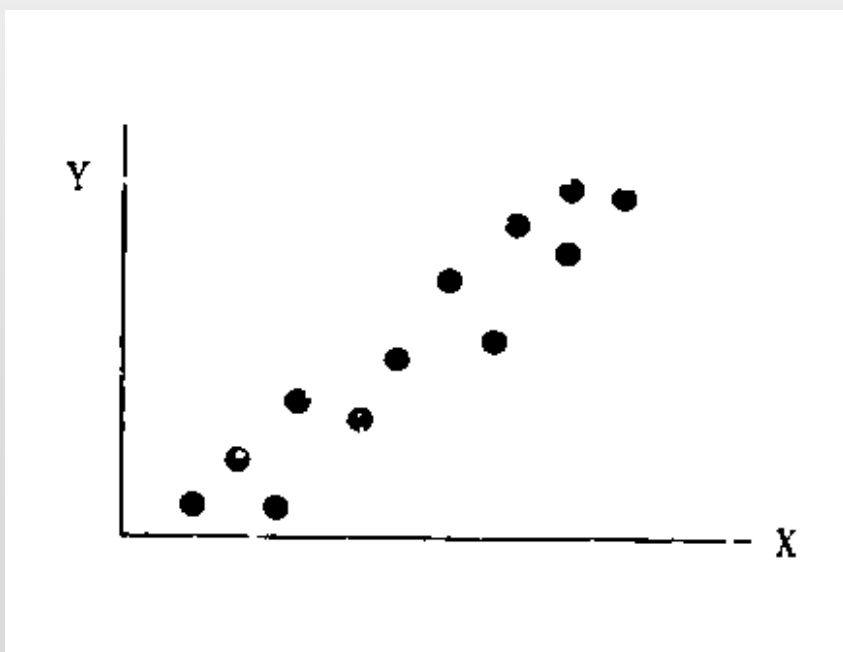
-1.00

มีความสัมพันธ์กันอย่างสมบูรณ์ทางลบ

การพิจารณาว่าตัวแปรทั้งสองมีความสัมพันธ์กันในทิศทางใด มีอยู่ 2 รูปแบบ คือ มีความสัมพันธ์กัน
ในทางบวก กับมีความสัมพันธ์กันในทางลบ



ความสัมพันธ์กันในทางบวก (Positive correlation) หมายความว่า ถ้าตัวแปรตัวหนึ่งมีค่าเพิ่มขึ้นค่าของตัวแปรอีกตัวหนึ่งก็จะเพิ่มขึ้นด้วย ในทางตรงข้าม ถ้าค่าของตัวแปรตัวหนึ่งลดลงค่าของตัวแปรอีกตัวหนึ่งก็จะลดลงด้วย





ความสัมพันธ์ในทางลบ (Negative correlation) มีความหมายตรงกันข้ามกับความสัมพันธ์ทางบวก คือ ถ้าค่าของตัวแปรหนึ่งเพิ่มขึ้น ค่าของตัวแปรอีกตัวหนึ่งจะลดลง หรือ ถ้าค่าของตัวแปรตัวหนึ่งลดลง ค่าของตัวแปรอีกตัวหนึ่งจะเพิ่มขึ้น





ในการศึกษาความสัมพันธ์ระหว่างตัวแปรสองตัว ค่าสัมประสิทธิ์สหสัมพันธ์ (Correlation coefficient) ที่ใช้แสดงปริมาณหรือขนาดของความสัมพันธ์ มีวิธีการหาที่แตกต่างกันไปตามลักษณะของข้อมูล กล่าวคือ ถ้าข้อมูลอยู่ในมาตราอันดับหรือมาตราอัตราส่วน ค่าสัมประสิทธิ์สหสัมพันธ์หาได้ด้วยวิธีการหาค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สันโปรดักโมเมนต์ (Pearson product moment correlation coefficient) ถ้าข้อมูลอยู่ในมาตราเรียงอันดับก็หาได้ด้วยวิธีการหาค่าสัมประสิทธิ์สหสัมพันธ์แบบสเปียร์แมนแรนจ์ (Spearman rank correlation coefficient)



สัมประสิทธิ์สหสัมพันธ์แบบสเปียร์แมนแรงค์

(Spearman rank correlation coefficient)

ในการวัดหรือเก็บข้อมูลบางครั้งต้องใช้วิธีการจัดอันดับที่ เช่น ผลการประกวดมารยาท การประกวดงานฝีมือ การคัดลายมือ เป็นต้น นิยมวัดโดยการจัดอันดับที่มากกว่าการให้เป็นคะแนนดิบ เมื่อต้องการหาความสัมพันธ์ระหว่างข้อมูลของตัวแปรที่เป็นอันดับที่จะต้องใช่วิธีการที่เรียกว่า “สหสัมพันธ์สเปียร์แมนแรงค์ (Spearman rank correlation)” ซึ่งมีสูตรในการคำนวณดังนี้

$$r_s = 1 - \frac{6 \sum d^2}{n(n^2 - 1)}$$

เมื่อ	r_s	แทน	ค่าสัมประสิทธิ์สหสัมพันธ์แบบสเปียร์แมนแรงค์
	d	แทน	ผลต่างของอันดับที่
	n	แทน	จำนวนคู่ของข้อมูล



ตัวอย่างที่ 2 จากตารางข้อมูลของธนาคารโลกปี 2513 ของประเทศขนาดกลางคือไม่เล็กเกินไปและโตเกินไปคือ ประชากรตั้งแต่ 12 ล้านไปถึงไม่เกิน 90 ล้าน ต้องการจะศึกษาถึงความสัมพันธ์ระหว่างอัตราเพิ่มประชากรกับ อัตราเพิ่มของรายได้ประชาชาติว่ามีความสัมพันธ์กันมากน้อยแค่ไหน การที่ไม่รวมเอาประเทศที่ใหญ่เกินไปหรือเล็กเกินไปมาพิจารณาด้วยเพราะเกรงว่าจะมีตัวแปรอื่น ๆ เข้ามาเกี่ยวข้อง



ประเทศ	ประชากร (ล้าน)	อัตราการเติบโตปี 2563				d	d ²
		ประชากร		รายได้ประชาชาติ			
		x	ลำดับ	y	ลำดับ		
บราซิล	88	3	2	1.6	10	8	64
ไนจีเรีย	62	2.4	4	-0.3	12.5	8.5	72.25
เยอรมันนีตะวันตก	60	1.0	11	3.4	5.5	5.5	30.25
สหราชอาณาจักร	55	0.7	14	2.0	8	6	36
อิตาลี	52	0.8	13	4.0	3	10	100
ฝรั่งเศส	49	1.1	9.5	3.7	4	5.5	30.25
เม็กซิโก	47	3.5	1	3.4	5.5	4.5	20.25
สเปน	32	0.9	12	6.5	1	11	121
อียิปต์	31	2.5	3	1.6	10	7	49
พม่า	26	2.1	6	1.6	10	4	16
ยูโกสลาเวีย	20	1.1	9.5	4.2	2	7.5	56.25
อัฟกานิสถาน	16	2.0	7	-0.3	12.5	5.5	30.25
เนเธอร์แลนด์	12	1.3	8	3.0	7	1	1
อัลจีเรีย	12	2.3	5	-3.5	14	9	81
รวม							$\sum d^2 = 707.5$



$$r_s = 1 - \frac{6 \sum d^2}{n(n^2 - 1)}$$

$$= 1 - \frac{6(707.5)}{14(14^2 - 1)}$$

$$= 1 - \frac{4245}{2744 - 14}$$

$$= 1 - \frac{4245}{2730} = 1 - 1.55494 = -0.55$$

จากการคำนวณพบว่า เมื่อจำนวนประชากรเพิ่ม รายได้ประชาชาติจะลดลง
ความสัมพันธ์ของตัวแปรทั้งสองจะเป็นไปในทางกลับกัน



สัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน (Pearson Correlation)

ถ้าข้อมูลที่ต้องการวัดความสัมพันธ์อยู่ในมาตราอันดับภาค หรือมาตราอัตราส่วน ค่าสัมประสิทธิ์สหสัมพันธ์จะหาได้ด้วยการหาค่าสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน เขียนแทนด้วยสัญลักษณ์ ρ แต่ในทางปฏิบัติจะหาค่า ρ ไม่ได้เนื่องจากเป็นการคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์จากตัวอย่างจึงต้องทำการประมาณค่าของ ρ ด้วย r โดยมีสูตรการคำนวณดังนี้

$$r_{xy} = \frac{\sum XY - n\bar{X}\bar{Y}}{\sqrt{(\sum X^2 - n\bar{X}^2)(\sum Y^2 - n\bar{Y}^2)}}$$



จากตัวอย่างที่ 1 ตารางแสดงข้อมูลเปอร์เซ็นต์เกษตรกรรมและค่าแตกต่างของรายได้

เปอร์เซ็นต์ เกษตรกรรม (X)	ค่าแตกต่างของ รายได้ (Y)	X^2	XY	Y^2
2.13	809	4.5369	1723.17	654,481
2.52	763	6.3504	1922.76	582,169
11.86	612	140.6596	7258.32	374,544
2.55	492	6.5025	1254.6	242,064
2.87	679	8.2369	1948.73	461,041
4.23	635	17.8929	2686.05	403,225



เปอร์เซ็นต์ เกษตรกร (X)	ค่าแตกต่างของ รายได้ (Y)	X^2	XY	Y^2
4.62	859	21.3444	3968.58	737,881
5.19	228	26.9361	1183.32	51,984
6.43	897	41.3449	5767.71	804,609
6.70	867	44.89	5808.9	751,689
1.53	513	2.3409	784.89	263,169
1.87	335	3.4969	626.45	112,225
10.38	868	107.7444	9009.84	753,424
62.88	8,557	432.2768	43943.32	6,192,505

จงหาค่าสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน



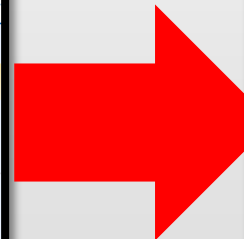
$$\begin{aligned}r_{xy} &= \frac{\sum XY - n\bar{X}\bar{Y}}{\sqrt{(\sum X^2 - n\bar{X}^2)(\sum Y^2 - n\bar{Y}^2)}} \\ &= \frac{43943.32 - 13(4.837)(658.231)}{\sqrt{(432.2768 - 13(4.837)^2)(6,192,505 - 13(658.231)^2)}} \\ &= 0.301\end{aligned}$$



ขั้นตอนที่ 2 Click Analyze → Regression → Linear จะได้หน้าจอดังรูป

The screenshot shows the IBM SPSS software interface. The 'Analyze' menu is open, and the 'Regression' option is selected. Under 'Regression', the 'Linear...' option is highlighted. In the background, a data table is visible with the following content:

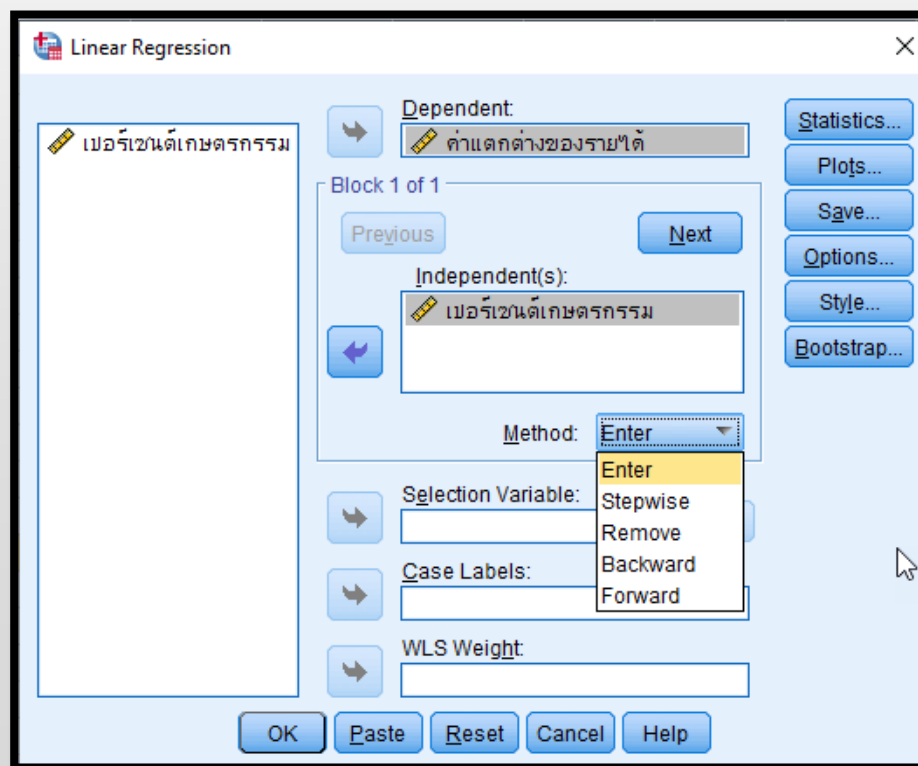
งรายใด	var	va
809		
763		
612		
492		
679		
635		



The screenshot shows the 'Linear Regression' dialog box. The 'Dependent' variable is 'ค่าแตกต่างของรายใด' (Difference in values). The 'Independent(s)' variable is 'เปอร์เซ็นต์เกษตรกร' (Percentage of farmers). The 'Method' is set to 'Enter'. The dialog box also includes buttons for 'Statistics...', 'Plots...', 'Save...', 'Options...', 'Style...', and 'Bootstrap...'. At the bottom, there are buttons for 'OK', 'Paste', 'Reset', 'Cancel', and 'Help'.



ขั้นตอนที่ 2 นำตัวแปรค่าแตกต่างของรายได้ใส่ไว้ที่ช่อง Dependent : และตัวแปรเปอร์เซ็นต์เกษตรกรรมใส่ไว้ที่ช่อง Independent(s) : ในส่วนของ Method ให้เลือก Enter ดังภาพประกอบ





ขั้นตอนที่ 3 คลิก OK จะได้ผลลัพธ์ดังนี้

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.301 ^a	.091	.008	215.137

a. Predictors: (Constant), เปอร์เซนต์เกษตรกรรม

จากตารางผลลัพธ์ พบว่า ค่าสัมประสิทธิ์สหสัมพันธ์ของเพียร์สันมีค่าเท่ากับ 0.301



Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	561.826	109.597		5.126	.000
	เปอร์เซ็นต์เกษตรกรรวม	19.931	19.006	.301	1.049	.317

a. Dependent Variable: ค่าแตกต่างของรายได้

จากตารางผลลัพธ์ พบว่า $a = 561.826$ และ $b = 19.931$

สามารถเขียนสมการถดถอยเชิงเส้นตรงดังนี้ $Y = 561.826 + 19.931X$



ตัวอย่างที่ 3 จากข้อมูลที่ได้จากการสำรวจของโรงแรมแห่งหนึ่งในประเทศไทยเกี่ยวกับ รายได้ต่อเดือนและราคาห้องพักที่นักท่องเที่ยวเข้าพักจำนวน 10 คน เป็นดังนี้

นักท่องเที่ยวคนที่	1	2	3	4	5	6	7	8	9	10
รายได้/เดือน	3,000	3,560	4,120	4,300	4,870	5,230	5,400	5,750	5,980	6,350
ราคาห้องพัก	225	200	459	445	439	577	722	903	1350	1360

ให้นักศึกษาวิเคราะห์ค่าสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน พร้อมทั้งหาสมการถดถอย



❖ การวิเคราะห์การถดถอยเชิงพหุคูณ (Multiple Linear Regression Analysis)

เป็นการศึกษาความสัมพันธ์ระหว่างตัวแปรต้น (X) และตัวแปรตาม (Y) ที่มีลักษณะเหมือนกันกับวิธี Simple Linear Regression คือ ความสัมพันธ์ระหว่างตัวแปรต้นและตัวแปรตามเป็นแบบเชิงเส้นตรง แต่ที่แตกต่าง คือ Multiple Linear Regression Analysis (Freeman, 2009) จะเป็นการศึกษาความสัมพันธ์ที่มีตัวแปรต้นมากกว่า 1 ตัว โดยความสัมพันธ์แทนด้วยสมการทางคณิตศาสตร์

$$Y = a + b_1X_1 + b_2X_2 + \dots + b_kX_k$$

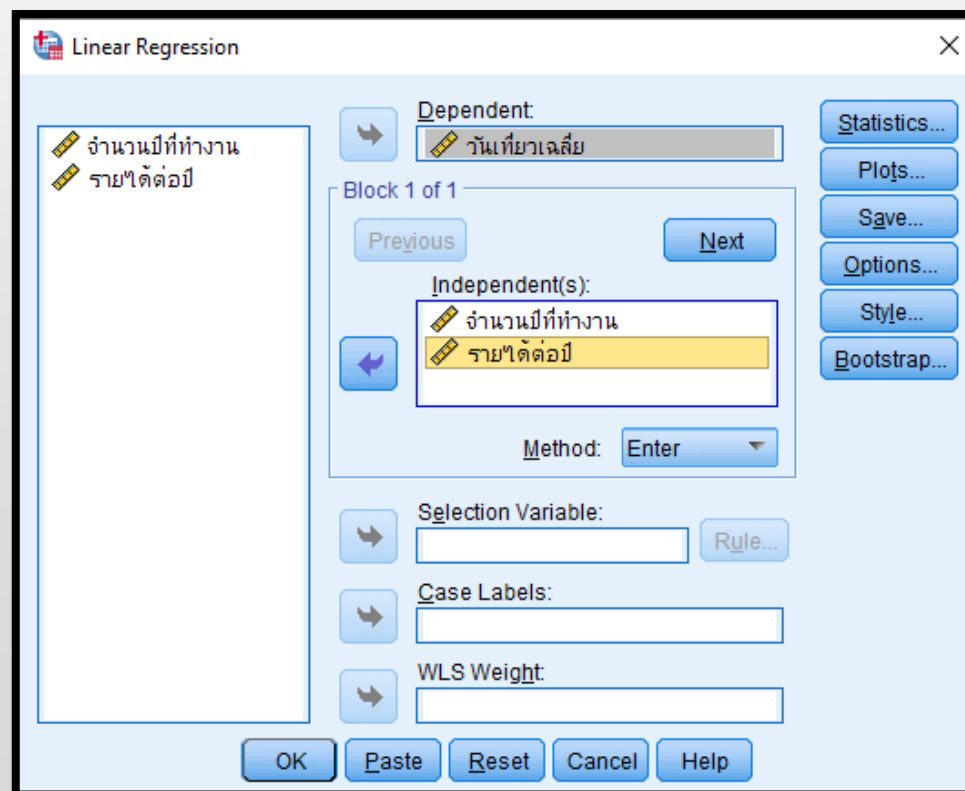


ตัวอย่างที่ 4 การศึกษาเพื่อพยากรณ์จำนวนวันเฉลี่ยที่นักท่องเที่ยวนำมาใช้ในการพักผ่อน ซึ่งมีความสัมพันธ์ กับจำนวนปีที่ทำงาน (Years of working) และรายได้ต่อปี (Yearly income) กำหนดตัวแปร ดังนี้ ตัวแปรตาม (Y) ได้แก่ จำนวนวันเฉลี่ยที่นักท่องเที่ยวนำมาใช้ในการท่องเที่ยว ตัวแปรต้น (X) ได้แก่ จำนวนปีที่ทำงาน แทนด้วย X_1 และรายได้ต่อปี แทนด้วย X_2

นักท่องเที่ยวดคนที่	1	2	3	4	5	6	7	8	9	10
จำนวนวันเฉลี่ยที่ ใช้เที่ยว	3	5	2	3	4	4	5	6	7	9
จำนวนปีที่ทำงาน	5	9	4	4	8	6	7	12	10	15
รายได้ต่อปี	100,000	150,000	80,000	90,000	150,000	120,000	200,000	180,000	85,000	250,000



ขั้นตอนที่ 2 นำตัวแปรวันเที่ยวเฉลี่ยใส่ไว้ที่ช่อง Dependent : และตัวแปรจำนวนปีที่ทำงาน และ รายได้ต่อปีใส่ไว้ที่ช่อง Independent(s) : ในส่วนของ Method ให้เลือก Enter ดังภาพประกอบ





ขั้นตอนที่ 3 คลิก OK จะได้ผลลัพธ์ดังนี้

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.948 ^a	.898	.869	.759

a. Predictors: (Constant), รายได้ต่อปี, จำนวนปีที่ทำงาน

จากตารางผลลัพธ์ พบว่า ค่าสัมประสิทธิ์สหสัมพันธ์ของเพียร์สันมีค่าเท่ากับ 0.948



Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.586	.683		.858	.420
	จำนวนปีที่ทำงาน	.517	.097	.916	5.331	.001
	รายได้ต่อปี	1.645E-6	.000	.044	.257	.804

a. Dependent Variable: วันเที่ยวเฉลี่ย

จากตารางผลลัพธ์ พบว่า $a = 0.586$ และ $b_1 = 0.517$, $b_2 = 0.00000016$

สามารถเขียนสมการถดถอยเชิงเส้นตรงดังนี้ $Y = 0.586 + 0.517 X_1 + 0.00000016 X_2$