

แผนบริหารการสอนประจำบทที่ 8

วิชา 4114601 การวิเคราะห์สถิติขั้นสูงด้วยโปรแกรมสำเร็จรูปทางสถิติ

แผนบริหารการสอน บทที่ 8 การวิเคราะห์การถดถอยที่มีความสัมพันธ์ในรูปแบบต่างๆ

เวลา 4 ชั่วโมง

สาระสำคัญ

ในบทนี้จะกล่าวถึงการวิเคราะห์การถดถอยที่มีความสัมพันธ์ในรูปแบบต่างๆซึ่งจะใช้ในกรณีที่มีความสัมพันธ์ระหว่างตัวแปรอิสระกับตัวแปรตามไม่เป็นเส้นตรงโดยจะแบ่งออกเป็น 2 ประเภทคือ เมื่อตัวแปรตามมีความสัมพันธ์กับตัวแปรอิสระในรูปแบบไม่เป็นเชิงเส้น และ ตัวแปรตามกับสัมประสิทธิ์ความถดถอยมีความสัมพันธ์ในรูปแบบไม่เชิงเส้น รวมทั้งจะกล่าวถึงการวิเคราะห์ความสัมพันธ์แบบโปรบิทโดยใช้โปรแกรมสำเร็จรูปทางสถิติในการวิเคราะห์

จุดประสงค์การเรียนรู้

นักศึกษาสามารถวิเคราะห์ความสัมพันธ์ของตัวแปรอิสระกับตัวแปรตามที่มีความสัมพันธ์รูปแบบต่างๆที่ไม่ใช่เส้นตรงได้อย่างถูกต้องและสามารถแปลความหมายผลการวิเคราะห์จากการใช้โปรแกรมสำเร็จรูปทางสถิติได้

กิจกรรมการเรียนการสอน

1. นำเสนอ PowerPoint เนื้อหาเกี่ยวกับการวิเคราะห์การถดถอยที่มีความสัมพันธ์ในรูปแบบต่างๆจากเอกสารประกอบการสอนการวิเคราะห์สถิติขั้นสูงด้วยโปรแกรมสำเร็จรูปทางสถิติ
2. ให้นักศึกษาทำแบบฝึกหัดท้ายบท

สื่อการเรียนรู้

1. PowerPoint
2. เอกสารประกอบการสอนรายวิชา 4114601 การวิเคราะห์สถิติขั้นสูงด้วยโปรแกรมสำเร็จรูปทางสถิติ

การวัดและประเมินผล

1. ประเมินผลจากแบบฝึกหัด
2. ประเมินผลจากการอภิปรายร่วมกันในชั้นเรียน
3. ประเมินผลจากการทดสอบย่อยรายบท

บทที่ 8

การวิเคราะห์การถดถอยที่มีความสัมพันธ์ในรูปแบบต่างๆ

(The relationship in different ways of Regression analysis)

การวิเคราะห์การถดถอยที่มีความสัมพันธ์ในรูปแบบต่างๆที่ไม่ใช่เชิงเส้นตรงนั้นมีหลายรูปแบบทั้งตัวแปรตามมีความสัมพันธ์กับตัวแปรอิสระในรูปแบบไม่เป็นเชิงเส้น และตัวแปรตามกับสัมประสิทธิ์ความถดถอยมีความสัมพันธ์ในรูปแบบไม่เป็นเชิงเส้น รวมทั้งความสัมพันธ์ที่เป็นแบบโปรบิท ซึ่งในบทนี้จะกล่าวถึงทั้ง 3 แบบดังนี้

1. การวิเคราะห์การถดถอยแบบไม่เป็นเชิงเส้น

1.1 รูปแบบของการวิเคราะห์การถดถอยที่ไม่ได้อยู่ในรูปเชิงเส้น (Nonlinear Regression) นั้นสามารถแบ่งประเภทของรูปแบบ เป็น 2 ประเภทคือ

1.1.1 ตัวแปรตามมีความสัมพันธ์กับตัวแปรอิสระในรูปแบบไม่เป็นเชิงเส้น เช่น $Y = \beta X^2 + E$ ซึ่งตัวแปรตามจะมีความสัมพันธ์กับตัวแปรอิสระในรูป quadratic แต่ยังสามารถประมาณค่าพารามิเตอร์ β โดยใช้เทคนิคการวิเคราะห์การถดถอยเชิงเส้นได้ เนื่องจากถ้าให้ $W = X^2$ สมการจะกลายเป็น $Y = \beta W + E$ ซึ่งอยู่ในรูปแบบของสมการเชิงเส้น

1.1.2 ตัวแปรตามกับสัมประสิทธิ์การถดถอยมีความสัมพันธ์ในรูปแบบที่ไม่เป็นเชิงเส้น เช่น $Y = e^{\beta x} + E$ ซึ่งเป็นรูปแบบของ exponential โดยในกรณีนี้สามารถทำการแปลง (Transform) ให้ความสัมพันธ์ของตัวแปรตามกับสัมประสิทธิ์การถดถอยมีความสัมพันธ์ในรูปแบบที่เป็นเชิงเส้นได้โดยทำการ take log

1.2 เงื่อนไขของการวิเคราะห์การถดถอยที่ไม่เป็นเชิงเส้น

1.2.1 ค่าความคลาดเคลื่อนต้องมีการแจกแจงแบบปกติ ($E \sim \text{Normal}$)

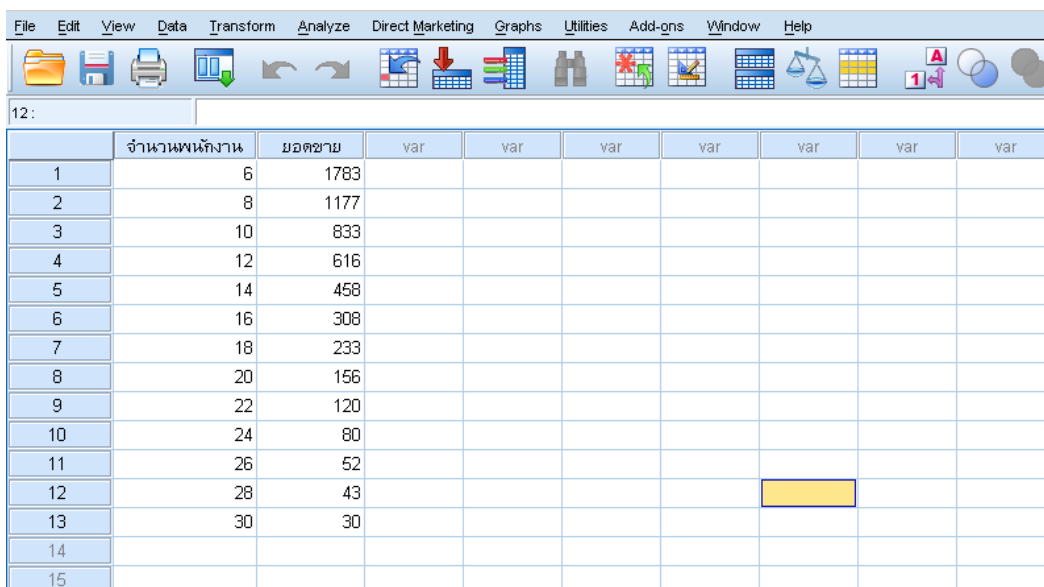
1.2.2 ค่าความแปรปรวนของความคลาดเคลื่อนต้องคงที่สำหรับแต่ละค่าของ X

1.3 ตัวอย่างการใช้โปรแกรมสำเร็จรูปทางสถิติวิเคราะห์ปัญหาการถดถอยที่ไม่เป็นเชิงเส้น จากข้อมูลจำนวนพนักงาน และยอดขายของร้านสะดวกซื้อแห่งหนึ่งเป็นดังนี้

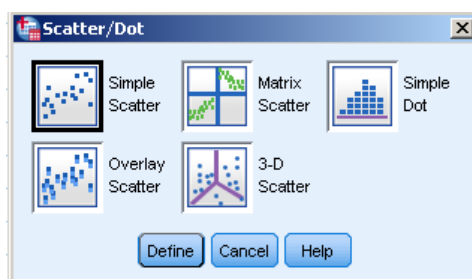
จำนวนพนักงาน	ยอดขาย (แสนบาทต่อเดือน)
6	1783
8	1177
10	833
12	616
14	458
16	30

จำนวนพนักงาน	ยอดขาย (แสนบาทต่อเดือน)
18	233
20	156
22	120
24	80
26	52
28	43
30	30

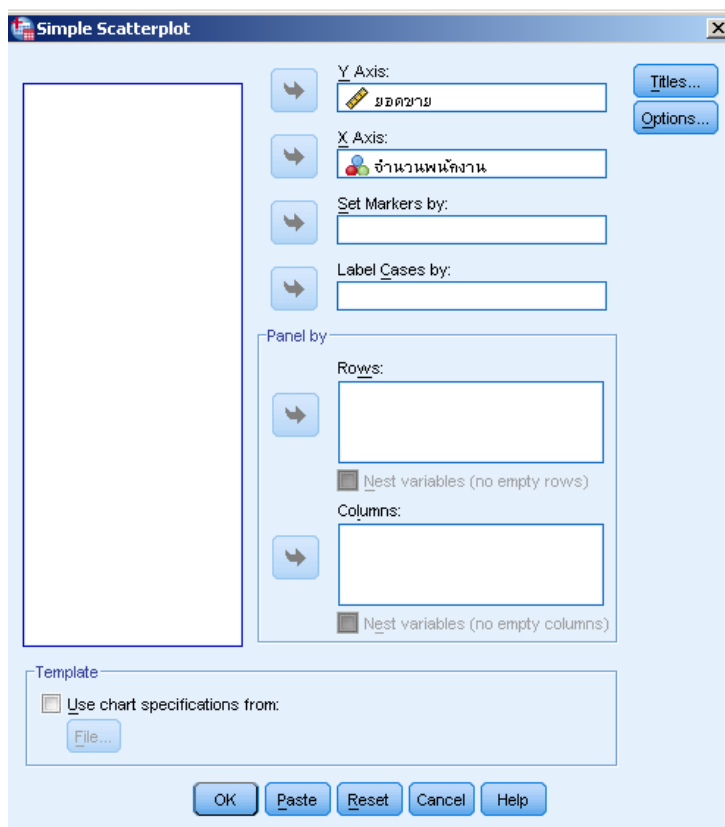
- สร้างข้อมูลดังตารางลงในโปรแกรมจากนั้นให้บันทึกข้อมูลโดยใช้ชื่อ file ว่า Nonlinear Regression ดังภาพ



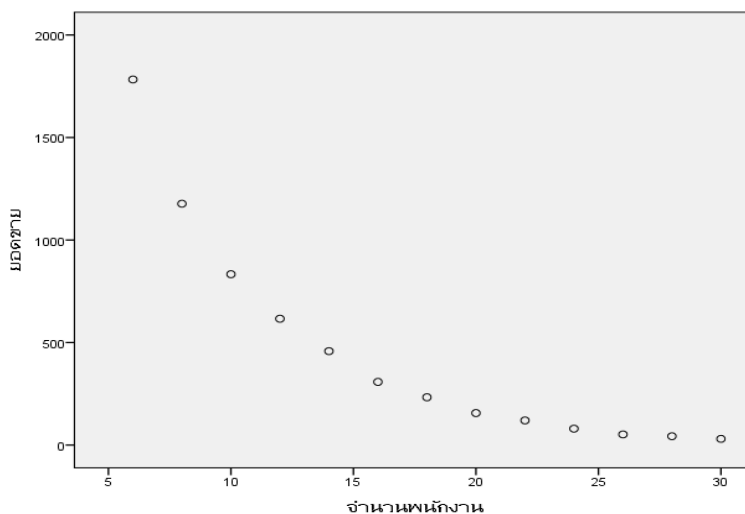
- ตรวจสอบความสัมพันธ์ของตัวแปรตามกับตัวแปรอิสระด้วยการ Plot Graph ดังนี้
Click → Graph → Legacy Dialogs → Scatter Plot/Dot เลือก Simple Scatter ดังภาพ



3. Click Define นำตัวแปรยอดขายใส่ไว้ที่ช่อง Y Axis และตัวแปรจำนวนพนักงานใส่ไว้ที่ช่อง X Axis ดังภาพ



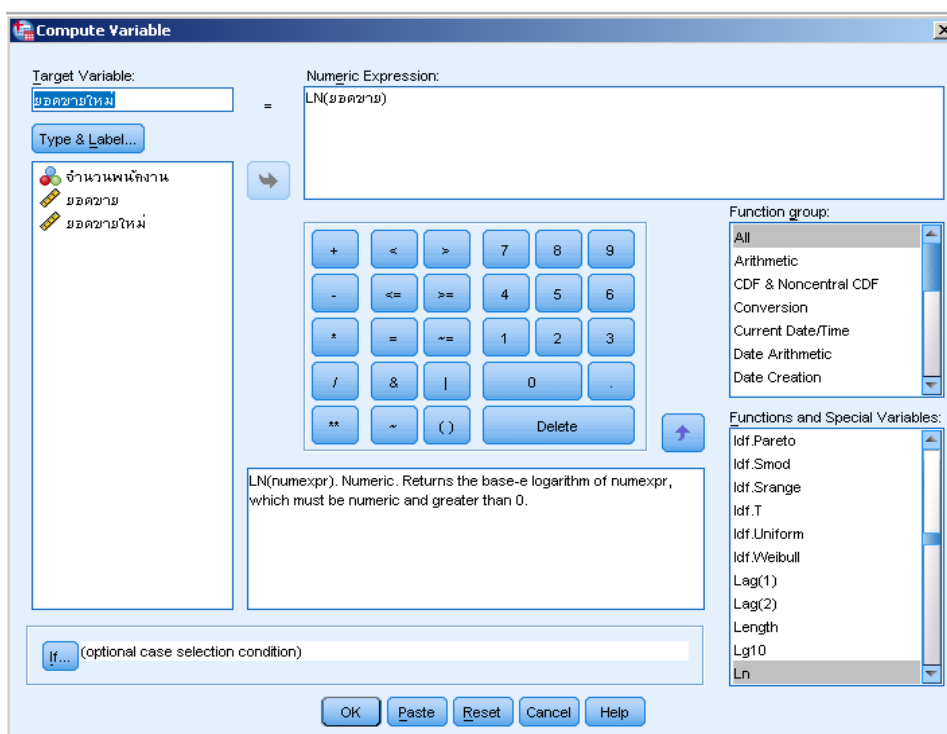
4. Click Ok จะได้กราฟ ดังภาพ



รูปที่ 8.1 รูปแสดงความสัมพันธ์ระหว่างตัวแปรยอดขายและจำนวนพนักงาน

จากรูปที่ 8.1 จะพบว่า ความสัมพันธ์ระหว่างตัวแปรยอดขายและจำนวนพนักงานมีความสัมพันธ์อยู่ในรูปที่ไม่เป็นเชิงเส้นโดยจะมีความสัมพันธ์อยู่ในรูปแบบของ exponential ดังนั้นต้องแปลงข้อมูลในที่นี่จะใช้การ Take Ln โดยจะใช้คำสั่ง SPSS ในการวิเคราะห์ดังนี้

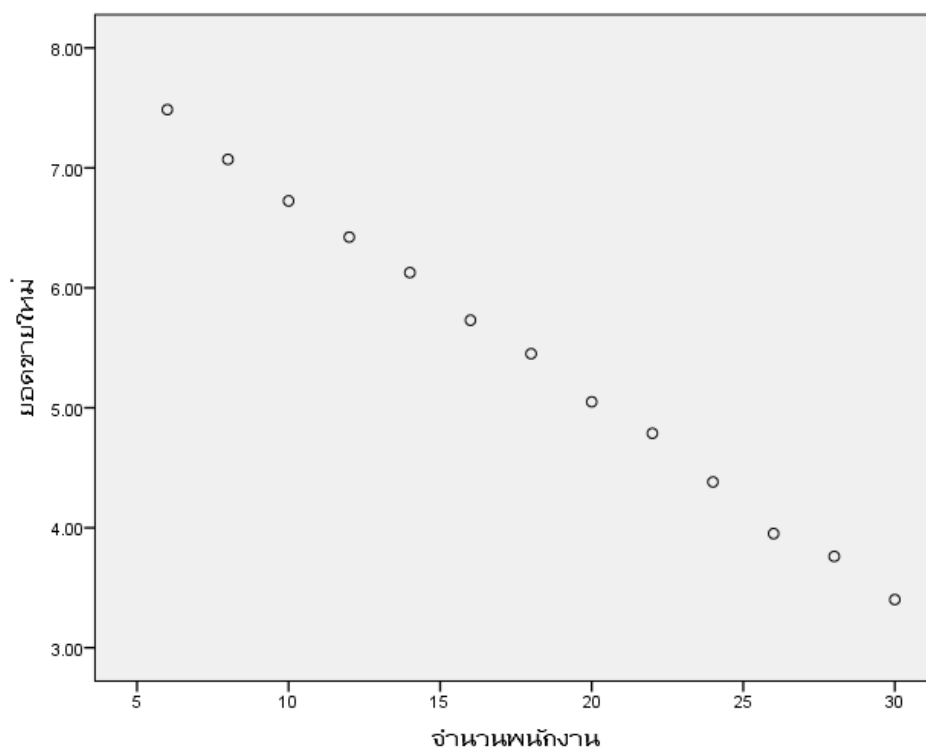
5. Click Analyze → Transform → Compute Variable ในส่วนของ Target Variable ใส่ชื่อตัวแปรตัวใหม่ที่ทำการแปลงข้อมูลในที่นี่ให้ชื่อว่าตัวแปรยอดขายใหม่ ในช่องของ Numeric Expression ใส่ LN(ยอดขาย)



6. Click OK จะได้ตัวแปรตัวใหม่ที่ทำการแปลงดังภาพ

	จำนวนพนักงาน	ยอดขาย	ยอดขายใหม่	var	var	var	var	var
1	6	1783	7.49					
2	8	1177	7.07					
3	10	833	6.73					
4	12	616	6.42					
5	14	458	6.13					
6	16	308	5.73					
7	18	233	5.45					
8	20	156	5.05					
9	22	120	4.79					
10	24	80	4.38					
11	26	52	3.95					
12	28	43	3.76					
13	30	30	3.40					
14								

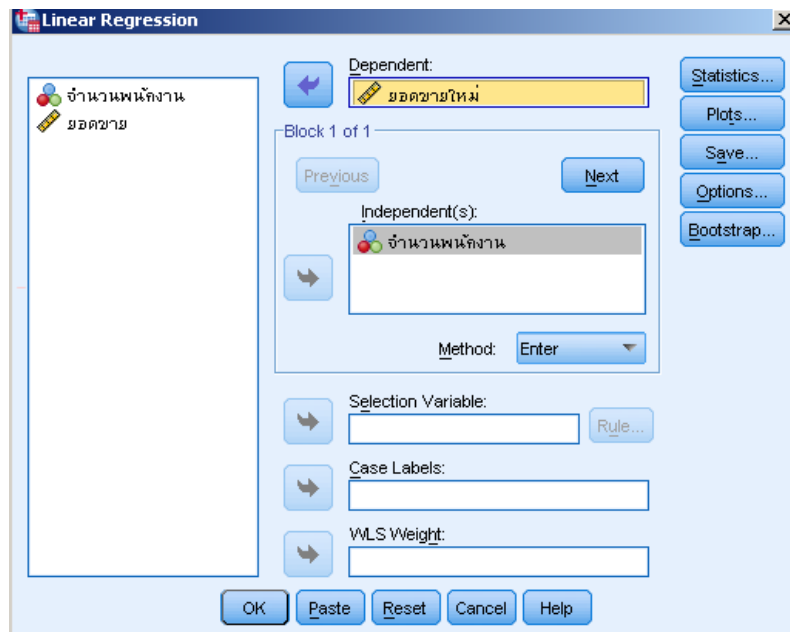
7. นำตัวแปรตัวใหม่ไปหาความสัมพันธ์กับตัวแปรอิสระด้วยการ Plot Graph จะได้ Graph ที่มีรูปแบบความสัมพันธ์เป็นเส้นตรงดังภาพ



รูปที่ 8.2 รูปแสดงความสัมพันธ์ระหว่างตัวแปรยอดขายใหม่และจำนวนพนักงาน

8. วิเคราะห์หาสมการถดถอยตามขั้นตอน ดังนี้

Click → Analyze → Regression → Linear นำตัวแปรยอดขายใหม่ใส่ไว้ที่ช่อง Dependent และนำตัวแปรจำนวนพนักงานใส่ไว้ที่ช่อง Independent(s) ในส่วนของ Method เลือก Enter ดังภาพ



9. Click OK จะได้ผลลัพธ์ดังตาราง

ตารางที่ 8.1 แสดงค่าสัมประสิทธิ์การถดถอย

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	8.459	.035		245.007	.000
	จำนวนพนักงาน	-.169	.002	-.999	-95.602	.000

a. Dependent Variable: ยอดขายใหม่

จากผลลัพธ์ในตารางที่ 8.1 จากโปรแกรม SPSS พบว่าสมการถดถอยคือ

$$\hat{Y} = 8.459 - 0.169X$$

แต่เนื่องจากค่า \hat{Y} เป็นค่าลอการิทึมดังนั้นจึงต้องแปลงกลับมาให้อยู่ในรูปข้อมูลเดิมโดย

$b_0 = e^{8.459} = 4717.338$ และ $b_1 = e^{-0.169} = 0.845$ ดังนั้นสมการถดถอยคือ

$$\hat{Y} = 4717.338 (0.845)^X$$

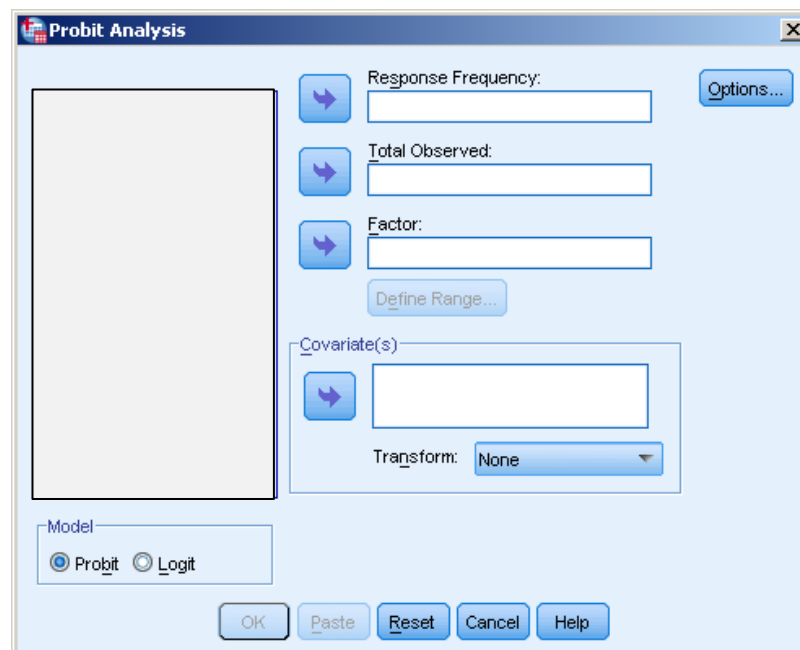
บทที่ 8 การวิเคราะห์การถดถอยที่มีความสัมพันธ์ในรูปแบบต่างๆ

2. การวิเคราะห์ความสัมพันธ์แบบโพรบิต

เทคนิคการวิเคราะห์ความสัมพันธ์แบบโพรบิตเป็นเทคนิคที่ใช้วัดความสัมพันธ์ของตัวแปรระหว่างตัวแปรระดับที่มีผลต่อการตอบสนอง หรือเป็นการศึกษาสัดส่วนของการตอบสนองที่มีต่อสิ่งกระตุ้น เช่น ต้องการวัดความสัมพันธ์ระหว่างส่วนลดของสินค้า กับสัดส่วนการซื้อสินค้าของลูกค้า ในที่นี้สิ่งกระตุ้น คือ ส่วนลดของสินค้า ส่วนการตอบสนอง คือการซื้อสินค้าของลูกค้า โดยที่แบบจำลองโพรบิต (Probit Model) และแบบจำลองโลจิท (Logit Model) เป็นแบบจำลองสำหรับตัวแปรตาม (dependent variable) ที่มีลักษณะเป็นตัวแปรเชิงคุณภาพ และมีค่าได้เพียง 2 ค่า (dichotomous variable) เช่น ใช่/ไม่ใช่ หรือไม่ใช่/ใช่ แต่ถ้การแจกแจงเป็นแบบ Logistic แบบจำลองจะเป็นแบบจำลองโลจิท และถ้การแจกแจงเป็นแบบปกติ (normal distribution) แบบจำลองจะเป็นแบบจำลองโพรบิต ด้วยเหตุผลทางทฤษฎีการศึกษาที่ได้จากแบบจำลองโลจิทและแบบจำลองโพรบิตจึงไม่แตกต่างกันเท่าใดนัก นอกจากขนาดตัวอย่างจะมากพอที่จะทำให้เห็นความแตกต่างที่ขวงปลายของการแจกแจง อย่างไรก็ตามเมื่อเปรียบเทียบวิธีการคำนวณเพื่อประมาณค่าสัมประสิทธิ์ตัวแปร แบบจำลองโลจิทจะมีวิธีการคำนวณที่ซับซ้อนน้อยกว่าแบบจำลองโพรบิต ดังนั้นการศึกษาวิถัยสวนใหญ่จึงนิยมใช้แบบจำลองโลจิทแทนแบบจำลองโพรบิต

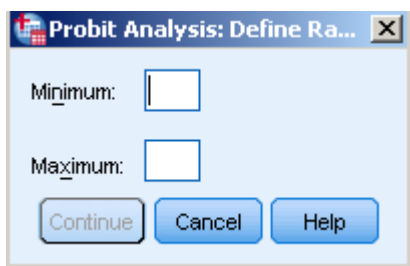
2.1 การใช้คำสั่งในโปรแกรมสำเร็จรูปทางสถิติในการวิเคราะห์ Probit

1. Click Analyze → Regression → Probit จะได้ดังภาพ

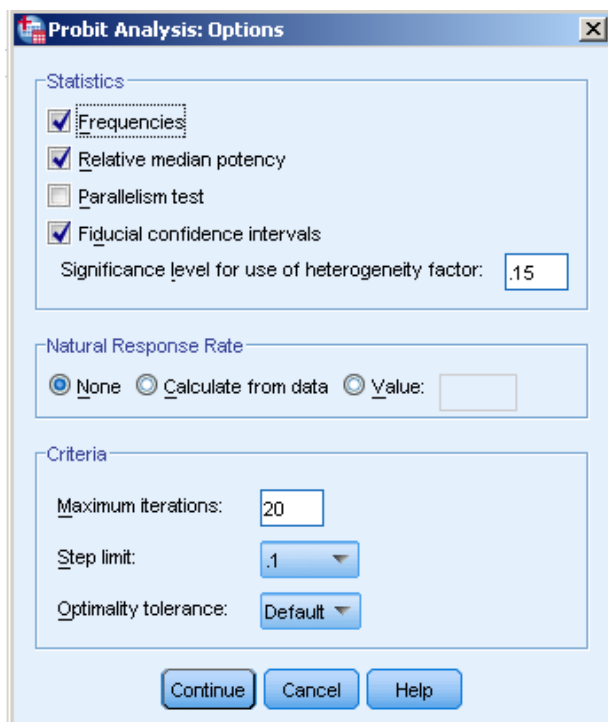


❖ ในส่วนของ Response frequency ใส่ตัวแปรที่เป็นตัวแปรตอบสนอง ซึ่งเป็นตัวแปรที่อยู่ในรูปของความถี่

- ❖ ในส่วนของ Total Observed ให้ใส่ตัวแปรที่บอกจำนวนตัวอย่างที่ถูกให้
สิ่งกระตุ้น
- ❖ ในส่วนของ Covariate(s) ให้ใส่ตัวแปรที่แสดงระดับของสิ่งกระตุ้น
- ❖ ในส่วนของ Factor จะเลือกตัวแปรใดในส่วนนี้หรือไม่ใส่ก็ได้ โดยตัวแปร
ที่จะใส่ในส่วนนี้ต้องเป็นตัวแปรเชิงกลุ่มที่จะต้องมีการกำหนดค่าของตัวแปรที่ต้องการวิเคราะห์ โดย
ต้องทำการคลิกปุ่ม Define Range จะได้ดังภาพ



2. Click Option จะได้ดังภาพ



- ❖ ในส่วนของ Statistics มีค่าให้เลือก 4 ค่า โดยถ้ามีตัวแปรใน Covariate มากกว่า 1 ตัวแปรจะไม่สามารถเลือก Relative median potency และ Fiducial confidence แต่จะสามารถเลือก Relative median potency และ Parallelism test โดยต้องเลือกตัวแปรใส่ในส่วนของ Factor ด้วย

- ❖ ในส่วนของ Natural Response Rate สามารถเลือกใช้ได้ 3 ค่าคือ
 - None ไม่ต้องคำนวณหา Response Rate
 - Calculate form data คำนวณหาอัตราตอบสนองจากข้อมูล
 - Value
- ❖ ในส่วนของ Criteria เป็นการระบุจำนวน Interaction ที่ต้องการให้มีการคำนวณ

3. บทสรุป

การนำตัวแบบที่เป็นเชิงเส้นมาใช้กับข้อมูลที่มีความสัมพันธ์รูปแบบอื่นๆ ที่ไม่เป็นเส้นตรง นั้นถือว่าไม่เหมาะสม อาจจะทำให้ความถูกต้องในการพยากรณ์ต่ำ ดังนั้นในการวิเคราะห์สมการถดถอยนั้นต้องมีการ Plot Graph เพื่อดูรูปแบบความสัมพันธ์เสียก่อนว่ามีความสัมพันธ์ในรูปแบบใด โดยถ้าหากข้อมูลใดสามารถแปลงให้อยู่ในรูปของเส้นตรงได้ควรแปลงข้อมูลก่อนการใช้ตัวแบบเชิงเส้น

แบบฝึกหัดบทที่ 8
เรื่อง การวิเคราะห์การถดถอยที่มีความสัมพันธ์ในรูปแบบต่างๆ

1. จากข้อมูลเนื้อที่ในการผลิตข้าวโพด และปริมาณผลผลิตดังตาราง

เนื้อที่ในการผลิตข้าวโพด (1,000 ไร่)	ผลผลิต (1,000 ตัน)
9008	535
7719	568
7802	588
7685	598
7317	590
6943	617
7040	619
6626	613
6040	633
5961	632

1.1 จงพล็อตกราฟเพื่อตรวจสอบรูปแบบความสัมพันธ์ของข้อมูล

1.2 จงแปลงข้อมูลเพื่อวิเคราะห์ถดถอยในรูปแบบความสัมพันธ์เชิงเส้น พร้อมทั้งอธิบายผลการวิเคราะห์ที่ได้

2. จากข้อมูลข้างล่าง

Y	1	1	2	2	3	3	4	4	5	5	6	6
X	44	40	56	62	74	76	82	78	68	74	54	60

จงพล็อตกราฟเพื่อตรวจสอบรูปแบบความสัมพันธ์ของข้อมูล

3. จงอธิบายความแตกต่างของการวิเคราะห์การถดถอยแบบโปรบิทและโลจิสติก

เอกสารอ้างอิงบทที่ 8

- กัลยา วานิชย์บัญชา. (2551). การวิเคราะห์สถิติขั้นสูงด้วย SPSS for Windows. พิมพ์ครั้งที่ 6. กรุงเทพฯ: โรงพิมพ์จุฬาลงกรณ์มหาวิทยาลัย, จุฬาลงกรณ์มหาวิทยาลัย.
- สุชาติ ประสิทธิ์รัฐสินธุ์. 2540. เทคนิคการวิเคราะห์ตัวแปรหลายตัวสำหรับการวิจัยทางสังคมศาสตร์และพฤติกรรมศาสตร์. พิมพ์ครั้งที่ 4. กรุงเทพฯ : เลียงเชียง.
- แบบจำลองถดถอยที่ไม่เป็นเชิงเส้น. (2557). ค้นเมื่อ 15 พฤษภาคม 2557, จาก http://lms.mju.ac.th/courses/159/locker/Econometrics2/content_con/9.htm .
- การถดถอยแบบไม่เปนเส้นตรง. (2557). ค้นเมื่อ 15 พฤษภาคม 2557, จาก <http://pirun.ku.ac.th/~faasatp/734415/data/chapter6.pdf>.