

สมการถดถอย (Regression)

สมการถดถอยเป็นการนำเอาข้อมูลที่เก็บไว้ในอดีตมาทำนายการเกิดขึ้นในอนาคตโดยใช้รูปแบบสมการเชิงเส้น นักศึกษาได้เรียนรู้เรื่องสหสัมพันธ์มาแล้วว่าเป็นการหาความสัมพันธ์ของตัวแปร 2 ตัว ในสหสัมพันธ์เราไม่สนว่าตัวแปรไหนเป็นตัวแปรตามตัวแปรไหนเป็นตัวแปรอิสระ แต่ในสมการถดถอยจะต้องระบุให้ชัดเจนว่าตัวแปรใดเป็นตัวแปรอิสระ (กำหนดเป็นค่า x) และตัวแปรใดเป็นตัวแปรตาม (กำหนดให้เป็น y) ตัวแปรอิสระเป็นเหตุและตัวแปรตามเป็นผลที่เกิดขึ้นเช่น เราทดลองใส่ปุ๋ยในนาข้าวในอัตราที่ไม่เท่ากันทำให้ข้าวแต่ละแปลงได้ผลผลิตไม่เท่ากัน ตัวแปรอิสระคืออัตราปุ๋ย ตัวแปรตามคือผลผลิตข้าว ตัวอย่างเช่นใส่ปุ๋ยให้ข้าวพันธุ์ IR8 4 อัตราแล้ววัดผลผลิตข้าวในแต่ละอัตราปุ๋ยได้ผลดังตารางที่ 1

ตารางที่ 1 ผลผลิตข้าว IR8 ที่ได้รับปุ๋ยไนโตรเจนในอัตราที่แตกต่างกัน 4 อัตรา

อัตราปุ๋ยไนโตรเจน (กก./ไร่)	ผลผลิต(กก./ไร่)
0	676.80
8	870.72
16	1065.76
24	1144.00

เมื่อนำข้อมูลมาเขียนกราฟใน Excel และเขียนสมการเชิงเส้นในรูปแบบ

$$y = bx + a$$

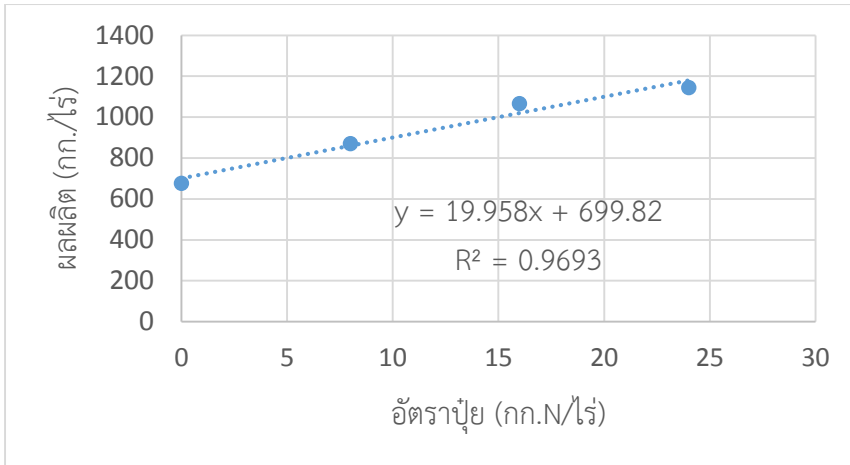
y คือตัวแปรตามในที่นี้คือผลผลิต

x คือตัวแปรอิสระในที่นี้คืออัตราปุ๋ยไนโตรเจน

b คือความชันเป็นค่าที่บอกว่าเมื่ออัตราปุ๋ยเปลี่ยนไป 1 กก./ไร่ผลผลิตจะเปลี่ยนไปกี่ กก./ไร่ ค่านี้เรียกว่าสัมประสิทธิ์การถดถอย (Regression coefficient)

a คือค่าคงที่หรือที่เรียกว่าจุดตัดแกน y เมื่อ x เท่ากับ 0

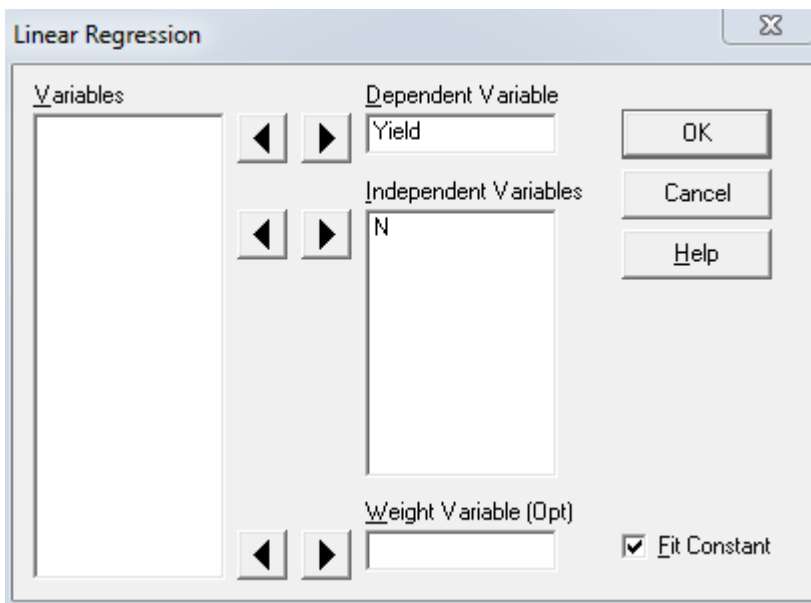
ใน Excel สามารถทำกราฟและคำนวณสมการให้ได้ดังภาพที่ 1 จะเห็นได้ว่า a มีค่าเท่ากับ 699.82 , b มีค่าเท่ากับ 19.958 ส่วนค่า R^2 ที่มีค่าเท่ากับ 0.9693 ก็หมายความว่า ค่า x ในสมการอธิบายอธิบายการเปลี่ยนแปลงของค่า y ได้ 96.9% แต่อย่างไรก็ตามยังต้องทดสอบทางสถิติก่อนว่าสมการนี้ยอมรับได้หรือไม่ โปรแกรม Excel คำนวณสมการให้ได้แต่ทดสอบการยอมรับทางสถิติไม่ได้ต้องไปคำนวณในโดยนำข้อมูลไปวางในโปรแกรม sx ดังภาพที่ 2 จากนั้นไปที่เมนู Statistics ต่อด้วย Linear Models ตามด้วย Linear Regression จะปรากฏหน้าต่างดังภาพที่ 3 ให้กำหนด ตัวแปรตาม (Dependent Variable) เป็นผลผลิต (Yield) และตัวแปรอิสระ (independent variable) เป็นอัตราปุ๋ย (N) จากนั้นกด OK จะได้ผลดังภาพที่ 4



ภาพที่ 1 กราฟและสมการถดถอยที่วาดและคำนวณด้วยโปรแกรม Excel

	N	Yield
1	0	676.8
2	8	870.72
3	16	1065.76
4	24	1144
*		

ภาพที่ 2 การวางข้อมูลในโปรแกรม sx โดย N คืออัตราปุ๋ย ไนโตรเจน Yield คือผลผลิตข้าว



ภาพที่ 3 การกำหนดตัวแปรตามและตัวแปรอิสระในการวิเคราะห์สมการถดถอยของโปรแกรม SX

Unweighted Least Squares Linear Regression of Yield					
Predictor Variables	Coefficient	Std Error	T	P	
Constant	699.824	37.6066	18.61	0.0029	
N	19.9580	2.51269	7.94	0.0155	
R-Squared	0.9693	Resid. Mean Square (MSE)		2020.36	
Adjusted R-Squared	0.9539	Standard Deviation		44.9484	
Source	DF	SS	MS	F	P
Regression	1	127463	127463	63.09	0.0155
Residual	2	4041	2020		
Total	3	131504			
Cases Included 4		Missing Cases 0			

ภาพที่ 4 ผลการทดสอบสถิติของสมการถดถอย

จากภาพที่ 4 อันดับแรกให้ดูตารางวิเคราะห์ความแปรปรวนด้านล่าง (ลูกขรชี้) ก่อนเป็นการทดสอบค่า F พิสูจน์ว่าความแปรปรวนที่เกิดขึ้นกับผลผลิต (ตัวแปรตาม) อธิบายได้ด้วยสมการถดถอยในสัดส่วนที่รับได้หรือไม่ ให้ดูที่ค่า P ถ้า P น้อยกว่า 0.05 แสดงว่าสมการใช้ได้ แต่ถ้ามากกว่า 0.05 แสดงว่าสมการใช้ไม่ได้ไม่ต้องดูอย่างอื่นต่อ

หากสมการใช้ได้ให้ดูต่อไปว่าค่า a และค่า b มีนัยสำคัญหรือไม่

ค่า b ถ้าใน column P ของค่า b น้อยกว่า 0.05 ก็แสดงว่าค่า b มีนัยสำคัญคงอยู่ในสมการ แต่หาก P มากกว่า 0.05 ก็ไม่สามารถใช้สมการนี้ได้ ในกรณีตัวอย่าง $p = 0.0155$ น้อยกว่า 0.05 ค่า b มีนัยสำคัญ คงอยู่ในสมการ สมการใช้ได้

สำหรับค่า a ก็ดูที่ column P เช่นกัน หาก p น้อยกว่า 0.05 ก็คงค่า a ไว้ในสมการ แต่หากว่า p มากกว่า 0.05 ก็ตัดค่า a ออกจากสมการแต่สมการก็ยังคงใช้ได้อยู่

ในกรณีนี้สุดท้ายก็ยังคงใช้สมการที่ Excel สร้างให้คือ

$$y = 19.958x + 699.82 \text{ (สมการ1)}$$

สมการนี้มีประโยชน์อย่างไร?

การใช้ประโยชน์

การใช้ประโยชน์ก็อย่างที่ได้อธิบายไปข้างต้นว่าใช้ทำนายเช่นเมื่อถดถอยเช่นสมการที่ 1 แล้วสามารถว่าหากใส่ปุ๋ยอัตราหนึ่งๆจะได้ผลผลิตข้าวเท่าไร เช่นต้องการทราบว่าหากใส่ปุ๋ยในอัตรา 20 กก./ไร่ จะได้ผลผลิตเท่าไร ก็แทนค่าในสมการให้ $x = 20$

$$y = 19.958x + 699.82$$

$$y = (19.958 \times 20) + 699.82$$

$$y = 1098.98 \text{ กิโลกรัมต่อไร่}$$

ข้อจำกัดที่สำคัญคือสมการถดถอยไม่สามารถทำนายค่าตัวแปรตามนอกเหนือขอบเขตของตัวแปรอิสระที่นำข้อมูลมาสร้างสมการได้ ตัวอย่างเช่น สมการ 1 ได้จากการทดลองใส่ปุ๋ย N ที่ต่างกัน 4 อัตรา ตั้งแต่ 0 – 24 กก./ไร่ ไม่สามารถทำนายได้ว่าการใส่ปุ๋ย 40 หรือ 50 กก./ไร่ จะได้ผลผลิตเท่าไร เพราะอัตราปุ๋ยสูงสุดที่นำมาสร้างสมการคือ 24 กก./ไร่เท่านั้น

แบบฝึกหัด

1. ร้านค้าแห่งหนึ่งได้กำหนดราคาขายกระเป่ายี่ห้อพร้อมบินตั้งแต่เดือนกรกฎาคมถึงธันวาคมเอาไว้แตกต่างกัน โดยบางเดือนก็ขายราคาเต็ม บางเดือนก็มีโปรโมชั่นลดราคาแล้วบันทึกยอดขายในแต่ละเดือนไว้ดังตาราง

เดือน	ราคาต่อหน่วย	ยอดขาย (ใบต่อเดือน)
กรกฎาคม	1,500	20
สิงหาคม	1,350	23
กันยายน	1,425	20
ตุลาคม	1,600	15
พฤศจิกายน	1,275	30
ธันวาคม	1,050	40

ให้สร้างสมการถดถอยระหว่างราคาและยอดขายโดยให้ราคาเป็นตัวแปรอิสระ พร้อมทั้งทดสอบว่าสมการที่ได้ยอมรับได้ทางสถิติหรือไม่ และทำนายว่า หากตั้งราคาขายไว้ที่ 1,200 บาทจะได้อยอดขายเท่าใด

2. ทดสอบอิทธิพลของอัตราการใช้ปุ๋ยฟอสฟอรัสต่อผลผลิตถั่วเหลืองโดยให้ปุ๋ยแตกต่างกัน 6 ระดับได้ผลผลิตดังตาราง

อัตราปุ๋ย P (กก./ไร่)	ผลผลิตถั่วเหลือง (กก./ไร่)
10	130
15	160
20	185
25	218
30	265
35	284

ให้วาดกราฟความสัมพันธ์ระหว่างอัตราปุ๋ยกับผลผลิต และหาสมการถดถอยโดยให้ อัตราปุ๋ยเป็นตัวแปรอิสระ และให้ผลผลิตเป็นตัวแปรตาม พร้อมทั้งทดสอบว่าสมการที่ได้ยอมรับได้ทางสถิติหรือไม่ และทำนายว่า หากใส่ปุ๋ย P ในอัตรา 5, 18 และ 27 กก.ต่อไร่จะได้ผลผลิตเท่าใด

3. บริษัทหนึ่งได้เก็บข้อมูลงบประมาณที่ใช้โฆษณา กับ ยอดขายสินค้าของบริษัทเป็นเวลา 5 ปีได้ผลดังตาราง

ปีที่	ค่าโฆษณา (ล้านบาท)	ยอดขาย (ล้านบาท)
1	0.35	10
2	0.80	20
3	0.60	15
4	1.00	30
5	0.83	22

ให้หาสมการถดถอยระหว่างงบประมาณกับยอดขาย พร้อมทั้งการทดสอบทางสถิติ และทำนายว่าหากตั้งงบประมาณในการโฆษณาไว้ 0.7 ล้านบาทจะได้ยอดขายเท่าใด