

# บทที่ 1

## บทนำ

การวิเคราะห์การถดถอย (regression analysis) เป็นวิธีทางสถิติที่ใช้หาความสัมพันธ์ระหว่างตัวแปรเชิงปริมาณตั้งแต่สองตัวขึ้นไปเพื่อใช้ในการพยากรณ์ค่าของตัวแปรหนึ่งจากตัวแปรอื่นๆ การวิเคราะห์การถดถอยใช้กันอย่างแพร่หลายในสาขาต่างๆ เช่น การแพทย์ วิทยาศาสตร์ ธุรกิจ เศรษฐศาสตร์และสังคมศาสตร์ เป็นต้น ตัวอย่างของการใช้สมการถดถอยในการวิเคราะห์ข้อมูลเช่น นักวิจัยต้องการทราบว่าปัจจัยใดบ้างที่มีผลต่อการนำระบบการจัดการความปลอดภัยมาใช้กับอุตสาหกรรมแปรรูปอาหารขนาดกลางและขนาดย่อมในประเทศไทย (พรเลิศ อภาณูทัต และ พรสิน สุภวาลย์, 2555) หรือการหาปัจจัยที่มีผลต่อการนำเทคโนโลยีสารสนเทศมาใช้ในอุตสาหกรรมขนาดกลางและขนาดย่อม (Ghobakhloo, Benitez-Amado & Arias-Aranda, 2011) หรือการหาปัจจัยที่มีผลต่อความน่าเชื่อถือในงานวิจัยทางการตลาด (Moorman, Deshpande & Zaltman, 1993) หรือปัจจัยที่มีผลต่อความพึงพอใจในการใช้ชีวิตของคนอเมริกัน (Sue et al, 1998) หรือเพื่อหาปัจจัยที่มีผลต่อความเครียดในการปฏิบัติงานของทหารในเขตสามจังหวัดชายแดนภาคใต้ (ไหมไทย ไชยพันธุ์และณัฐสุดา เต้พันธ์, 2555) เป็นต้น

ตัวแปรที่ใช้ในการวิเคราะห์การถดถอยมีสองแบบใหญ่ๆ คือ ตัวแปรอิสระ (independent variable หรือ regressor หรือ predictor) กับตัวแปรตาม (dependent variable) โดยที่ตัวแปรอิสระนั้นเป็นตัวแปร ที่ทราบค่าและนักวิจัยเป็นผู้กำหนดค่าของตัวแปรอิสระในการทดลองหรือการศึกษา โดยที่นักวิจัยคาดว่าตัวแปรอิสระนี้จะมีผลต่อตัวแปรตาม ซึ่งเป็นตัวแปรที่ไม่ทราบค่าและต้องการพยากรณ์หรือต้องการหาความสัมพันธ์ เช่น ในกรณีของตัวอย่างการวิจัยตลาดนั้นตัวแปรอิสระคือปัจจัยต่างๆ ที่นักวิจัยตลาดคาดว่าจะมีผลทำให้ยอดขายเปลี่ยนแปลงไปในขณะที่ตัวแปรตามคือมูลค่าของยอดขาย เป็นต้น ตัวแปรอิสระมักแทนด้วย  $X$  และตัวแปรตามมักแทนด้วย  $Y$

การวิเคราะห์การถดถอยระหว่างตัวแปรอิสระหนึ่งตัวและตัวแปรตามหนึ่งตัว โดยที่ตัวแปรทั้งสองมีความสัมพันธ์เชิงเส้นตรงกันเรียก การวิเคราะห์การถดถอยเชิงเส้นอย่างง่าย (simple linear regression analysis) หากความสัมพันธ์ระหว่างตัวแปรที่ไม่เป็นเส้นตรง เรียก nonlinear regression analysis หากมีตัวแปรอิสระมากกว่าหนึ่งตัวแต่ตัวแปรตามเพียงตัวเดียวเรียก การวิเคราะห์การ

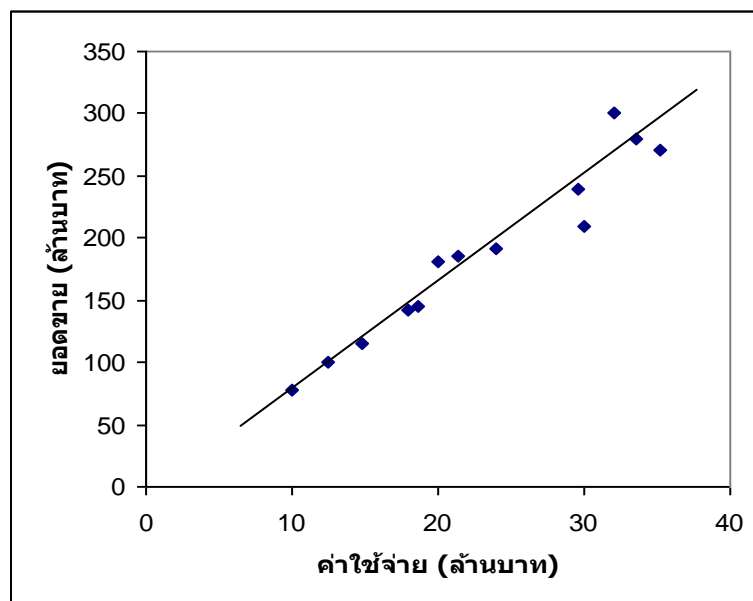
ถดถอยพหุ (multiple regression analysis) หากมีตัวแปรตามมากกว่า 1 ตัวแล้วการวิเคราะห์การถดถอยนั้นจะเรียกว่า multivariate regression analysis

วิธีการตรวจสอบความสัมพันธ์ระหว่างคู่ของตัวแปรอย่างง่ายทำได้โดยการสร้างแผนภาพกระจาย (scatter diagram) ระหว่างตัวแปรทั้งสองเพื่อดูว่าตัวแปรทั้งสองมีความสัมพันธ์กันหรือไม่อย่างไร ข้อดีของวิธีนี้คือ ง่ายและรวดเร็ว สามารถทำความเข้าใจได้ง่าย ผู้อ่านไม่จำเป็นต้องมีความรู้ทางสถิติเพื่อตีความ ข้อเสียของวิธีนี้คือ ทำการตรวจสอบความสัมพันธ์ได้ที่ละคู่ของตัวแปรและไม่มีค่าสถิติรองรับดังนั้นการแปลผลของแต่ละคนอาจแตกต่างกันได้โดยเฉพาะอย่างยิ่งในกรณีที่รูปที่ได้ไม่มีความชัดเจนพอหรือไม่มีแนวโน้ม

**ตัวอย่างที่ 1.1** การหาความสัมพันธ์ระหว่างค่าใช้จ่ายในการโฆษณาและยอดขายของสินค้าชนิดหนึ่งโดยใช้แผนภาพกระจาย โดยมีข้อมูลดังนี้

| ค่าใช้จ่าย (ล้านบาท) | ยอดขาย (ล้านบาท) |
|----------------------|------------------|
| 10.01                | 77.60            |
| 14.75                | 114.90           |
| 17.94                | 141.40           |
| 23.93                | 190.80           |
| 29.61                | 239.90           |
| 35.18                | 270.00           |
| 33.50                | 280.00           |
| 12.45                | 100.50           |
| 20.04                | 180.40           |
| 18.60                | 145.00           |
| 30.00                | 210.00           |
| 32.00                | 300.00           |
| 21.40                | 185.00           |

เนื่องจากนักวิจัยตลาดเชื่อว่าค่าใช้จ่ายในการโฆษณามีผลต่อยอดขายที่เปลี่ยนแปลงไป ดังนั้นค่าใช้จ่ายในการโฆษณาเป็นตัวแปรอิสระ ( $X$ ) และยอดขายเป็นตัวแปรตาม ( $Y$ ) โดยสามารถสร้างแผนภาพกระจายได้ดังภาพที่ 1.1



ภาพที่ 1.1 แผนภาพกระจายแสดงความสัมพันธ์ระหว่างค่าใช้จ่าย (ล้านบาท) และ ยอดขาย (ล้านบาท)

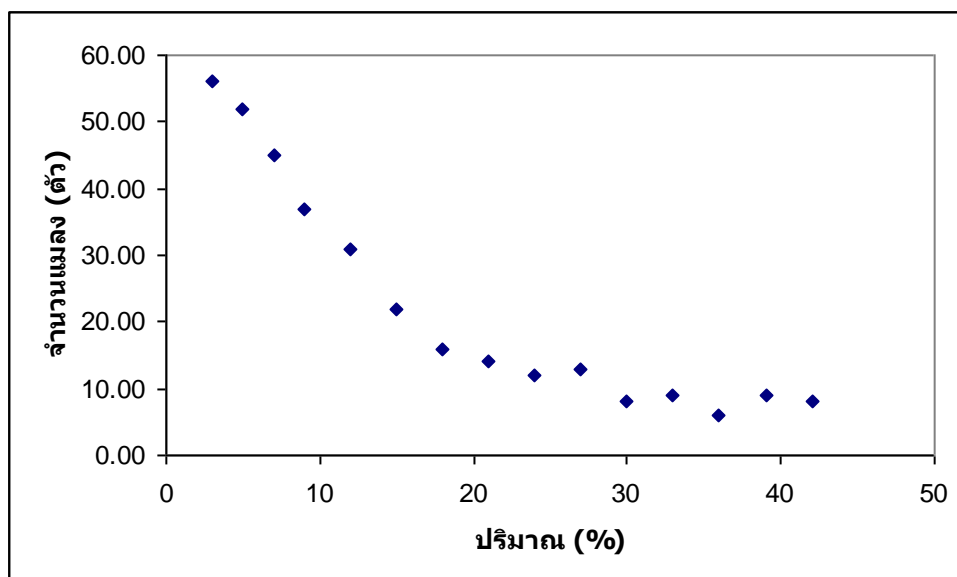
จะเห็นว่าค่าใช้จ่ายในการโฆษณาและยอดขายมีความสัมพันธ์เชิงเส้นตรงกัน โดยที่เมื่อค่าใช้จ่ายในการโฆษณาเพิ่มขึ้นทำให้ยอดขายเพิ่มมากขึ้นตามไปด้วยแต่อย่างไรก็ตามความสัมพันธ์ดังกล่าวไม่ได้เป็นไปอย่างสมบูรณ์เนื่องจากค่าของข้อมูลทั้งหมดไม่ได้อยู่บนเส้นตรง การที่ข้อมูลบางส่วนไม่อยู่บนเส้นตรงนั้นสาเหตุส่วนหนึ่งมาจากความแปรผันที่อาจเกิดขึ้นอย่างสุ่มที่ไม่สามารถอธิบายได้

บางครั้งความสัมพันธ์ระหว่างตัวแปรอาจเป็นเส้นโค้งในลักษณะต่าง ๆ กันเช่น โค้งพาราโบลา เอกซ์โปเนนเชียล รูปตัวเอส โค้งสองยอดหรือโค้งหลายยอด เป็นต้น หรือบางครั้งตัวแปรทั้งสองอาจไม่มีความสัมพันธ์กันเลยก็ได้ ในกรณีที่ตัวแปรทั้งสองไม่มีความสัมพันธ์ระหว่างกันแล้วไม่สามารถใช้การวิเคราะห์การถดถอยในการพยากรณ์ได้

ตัวอย่างที่ 1.2 การหาความสัมพันธ์ระหว่างปริมาณยาที่ใช้ในการทำละลายแมลง (%) และจำนวนแมลงที่ฟักเป็นตัว (ตัว) โดยใช้แผนภาพกระจายโดยมีข้อมูลดังนี้

| ปริมาณ (%) | จำนวนแมลง (ตัว) |
|------------|-----------------|
| 3          | 56.00           |
| 5          | 52.00           |
| 7          | 45.00           |
| 9          | 37.00           |
| 12         | 31.00           |
| 15         | 22.00           |
| 18         | 16.00           |
| 21         | 14.00           |
| 24         | 12.00           |
| 27         | 13.00           |
| 30         | 8.00            |
| 33         | 9.00            |
| 36         | 6.00            |
| 39         | 9.00            |
| 42         | 8.00            |

เนื่องจากนักวิจัยเชื่อว่าปริมาณยาที่ใช้มีผลต่อจำนวนแมลงที่ฟักเป็นตัวดังนั้นปริมาณยาเป็นตัวแปรอิสระ ( $X$ ) และจำนวนไข่แมลงที่ฟักเป็นตัวเป็นตัวแปรตาม ( $Y$ ) โดยสามารถสร้างแผนภาพกระจายได้ดังภาพที่ 1.2



ภาพที่ 1.2 แผนภาพกระจายแสดงความสัมพันธ์ระหว่างปริมาณยาที่ใช้ (%) และจำนวนแมลงที่ปักเป็นต้น (ตัว)

จะเห็นว่าความสัมพันธ์ระหว่างปริมาณยาที่ใช้และจำนวนไขแมลงที่ปักเป็นต้นไม่เป็นเส้นตรงแต่เป็นเอ็กซ์โปเนนเชียลที่แปรผกผันกัน โดยที่ปริมาณยาค่าๆ จำนวนไขแมลงที่ปักเป็นต้นจะมีจำนวนมาก แต่เมื่อเพิ่มปริมาณยามากขึ้นเพียงเล็กน้อยจำนวนไขแมลงที่ปักเป็นต้นจะลดลงจำนวนมากแบบทวีคูณ

กรณีที่ตัวแปรมีความสัมพันธ์กันในรูปแบบใดรูปแบบหนึ่งสามารถเขียนความสัมพันธ์ระหว่างตัวแปรในรูปของฟังก์ชัน  $Y = f(X)$  ได้เช่น กรณีของค่าใช้จ่ายในการโฆษณาที่บอกขายมีความสัมพันธ์กันเชิงเส้นตรงสามารถเขียนเป็นสมการเส้นตรงได้ดังนี้

$$Y = \beta_0 + \beta_1 X + \varepsilon \quad (1.1)$$

โดยที่  $\beta_0$  เป็นจุดตัดแกน  $Y$ ,  $\beta_1$  เป็นความชันหรืออัตราการเปลี่ยนแปลงของตัวแปร  $Y$  เมื่อ  $X$  เปลี่ยนแปลงไป 1 หน่วย ทั้ง  $\beta_0$  และ  $\beta_1$  เป็นค่าคงที่ซึ่งเรียกว่าพารามิเตอร์ (parameter) จากการที่จุดทุกจุดไม่อยู่บนเส้นตรงในแผนภาพกระจายเนื่องจากมีความแปรผันอย่างสุ่มหรือความคลาดเคลื่อน (random error) เกิดขึ้น ดังนั้นจึงต้องมีการเพิ่มพจน์ของ  $\varepsilon$  ขึ้นโดยที่ค่าความคลาดเคลื่อนนี้ถือเป็นตัวแปรสุ่มที่เกิดขึ้นจากตัวแปรอื่นๆ ที่อาจมีผลต่อตัวแปรตามแต่นักวิจัยไม่ได้นำมาพิจารณา สมการดังกล่าวเรียกว่าสมการถดถอยเชิงเส้นหรือตัวแบบการถดถอยเชิงเส้น (linear regression model) หากมีตัวแปรอิสระ  $k$  ตัวแล้วตัวแบบการถดถอยจะเป็นดังนี้

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon \quad (1.2)$$

โดยที่  $\beta_1$  เป็นอัตราการเปลี่ยนแปลงของตัวแปร  $Y$  เมื่อตัวแปรอิสระ  $X_1$  เปลี่ยนแปลงไป 1 หน่วย  $\beta_2$  เป็นอัตราการเปลี่ยนแปลงของตัวแปร  $Y$  เมื่อตัวแปรอิสระ  $X_2$  เปลี่ยนแปลงไป 1 หน่วยและ  $\beta_k$  เป็นอัตราการเปลี่ยนแปลงของตัวแปร  $Y$  เมื่อตัวแปรอิสระ  $X_k$  เปลี่ยนแปลงไป 1 หน่วย ตัวแบบถดถอยดังกล่าวนี้เรียกว่าตัวแบบการถดถอยเชิงเส้นพหุ (multiple linear regression model) หากตัวแปรอิสระและตัวแปรตามมีความสัมพันธ์ที่ไม่เป็นเชิงเส้นตรงกันจะเรียกตัวแบบนั้นว่าตัวแบบการถดถอยที่ไม่ใช่เชิงเส้น (nonlinear regression model หรือ polynomial regression model) ตัวอย่างของตัวแบบที่ไม่เป็นเส้นตรงนี้เช่นในตัวอย่างที่ 1.2 ซึ่งสามารถเขียนตัวแบบได้เป็น

$$Y = \beta_0 e^{\beta_1 X} + \varepsilon \quad (1.3)$$

ในกรณีของข้อมูลที่มีความสัมพันธ์ที่ไม่เป็นเส้นตรงกันนั้นสามารถแปลง (transform) ข้อมูลเพื่อให้ความสัมพันธ์ระหว่างตัวแปรอิสระและตัวแปรตามเป็นเส้นตรงแล้ววิเคราะห์โดยใช้ตัวแบบเชิงเส้นได้

ในการวิเคราะห์การถดถอยจำเป็นต้องมีข้อตกลง (assumption) ในการวิเคราะห์ซึ่งจำนวนข้อตกลงจะแตกต่างกันไปขึ้นอยู่กับประเภทของตัวแบบการถดถอย ข้อตกลงที่สำคัญเช่น การแจกแจงแบบปกติของตัวแปรตาม (normality) ความเป็นอิสระของตัวแปรตาม (independency) และความแปรปรวนของตัวแปรตามที่คงที่ (homoscedasticity) เป็นต้น บางครั้งพบว่านักวิจัยไม่ได้ความสำคัญกับการตรวจสอบข้อมูลว่าเป็นไปตามข้อตกลงที่กำหนดไว้หรือไม่ หากข้อมูลไม่เป็นไปตามข้อตกลงที่กำหนดแล้วอาจทำให้อำนาจการทดสอบ (power of test) ต่ำลงหรือความผิดพลาดในการพยากรณ์มากขึ้น นอกเหนือจากการละเมิดข้อตกลงแล้วการที่ข้อมูลบางค่ามีค่าที่ผิดปกติไปจากข้อมูลอื่นๆ (outlier) หรือการที่ตัวแปรอิสระบางตัวมีความสัมพันธ์กันเอง (multicollinearity) จะมีผลต่อคุณภาพของตัวแบบการถดถอยหรืออาจทำให้สร้างตัวแบบการถดถอยที่ไม่ถูกต้องอีกด้วย

งานวิจัยส่วนใหญ่จะมีตัวแปรอิสระที่นักวิจัยสนใจจำนวนมากแต่การคัดเลือกตัวแปรอิสระที่มีความจำเป็นจริงๆต่อตัวแปรตามนั้นมีด้วยกันหลายวิธีแต่ละวิธีอาจให้ผลลัพธ์ที่แตกต่างกันออกไปเนื่องจากเกณฑ์การพิจารณาที่ใช้แตกต่างกันเช่น การลดตัวแปรอิสระ (backward elimination method) การเพิ่มตัวแปรอิสระ (forward selection method) การเพิ่มตัวแปรอิสระแบบขั้นตอน (stepwise method) และการเลือกชุดที่ดีที่สุด (best subset method) เป็นต้น

การวิเคราะห์การถดถอยบางครั้งสามารถคำนวณได้ด้วยการใช้เครื่องคิดเลขซึ่งอาจใช้เวลานานและยุ่งยาก ปัจจุบันมีโปรแกรมทางสถิติจำนวนมากที่สามารถใช้วิเคราะห์การถดถอยได้ในหลายรูปแบบทั้งที่เป็นการคำนวณสำเร็จรูปและการเขียนคำสั่งเพื่อให้สอดคล้องกับความต้องการในการวิเคราะห์ขั้นสูง ตัวอย่างของโปรแกรมสำเร็จรูปทางสถิติ เช่น BMDP, MINITAB,

SAS, SPSS, SYSTAT และ JMP เป็นต้น สำหรับในหนังสือเล่มนี้จะใช้โปรแกรม MINITAB ในการคำนวณเนื่องจากเป็นโปรแกรมที่ใช้กันอย่างแพร่หลายในงานวิจัยทางวิทยาศาสตร์และง่ายแก่การใช้งาน

## สรุป

การหาความสัมพันธ์ระหว่างตัวแปรสองตัวอย่างง่ายสามารถทำได้โดยการสร้างแผนภาพกระจายระหว่างตัวแปรทั้งสองแต่เมื่อตัวแปรที่มีจำนวนมากขึ้นการสร้างแผนภาพกระจายอาจเป็นไปได้ นักสถิติจึงนำเอาการวิเคราะห์การถดถอยเข้ามาช่วยในการหาความสัมพันธ์ระหว่างตัวแปร นอกจากนี้การวิเคราะห์การถดถอยยังนำใช้ช่วยในการคัดเลือกตัวแปรอิสระที่มีผลต่อตัวแปรตาม

## คำถามท้ายบท

- 1.1 การตรวจสอบรูปแบบความสัมพันธ์ระหว่างตัวแปรสองตัวแปรอย่างง่ายสามารถใช้วิธีใดบ้าง
  - 1.2 จงวาดแผนภาพกระจายระหว่างตัวแปรทั้งสองแล้วอธิบายว่าตัวแปรทั้งสองมีความสัมพันธ์กันอย่างไร
- |    |    |    |    |    |    |    |    |    |
|----|----|----|----|----|----|----|----|----|
| X: | 21 | 13 | 20 | 25 | 19 | 24 | 16 | 13 |
| Y: | 13 | 6  | 12 | 7  | 19 | 10 | 24 | 19 |
- 1.3 ตัวแปรอิสระและตัวแปรตามคืออะไรพร้อมทั้งยกตัวอย่างประกอบ
  - 1.4 จงยกตัวอย่างการใช้สมการถดถอยเชิงเส้นพหุในการวิเคราะห์ข้อมูล
  - 1.5 นักการศึกษาต้องการทราบว่าคะแนนสอบเข้ามหาวิทยาลัยมาจากสาเหตุใดบ้าง เช่น IQ ของนักเรียน ฐานะทางครอบครัว เกรดเฉลี่ยสะสมในมัธยม 6 เป็นต้น จงระบุว่าตัวแปรใดเป็นตัวแปรอิสระและตัวแปรใดเป็นตัวแปรตาม
  - 1.6 จงยกตัวอย่างของข้อมูลที่มีความสัมพันธ์แบบไม่เป็นเส้นตรง
  - 1.7 จงวาดกราฟระหว่างจุดต่อไปนี้
    - (1) (0,2) กับ (2,6)
    - (2) (0,4) กับ (2,6)
    - (3) (0,-2) กับ (-1,-6)
    - (4) (0,-4) กับ (3,-7)

1.8 จงวาดกราฟของสมการดังต่อไปนี้

(1)  $Y = 3 + 2X$

(2)  $Y = X$

(3)  $Y = 4 - 5X$

(4)  $Y = 1 + 3X$

1.9 จากกราฟในข้อ 1.8 จงหาจุดตัดแกน  $Y$  และความชัน

1.10 จงอธิบายความหมายของ  $\beta_0$  และ  $\beta_1$