บทที่ 3
การวิเคราะห์การถดถอย
(Multiple Regression Analysis)

การวิเคราะห์การถดถอยและสหสัมพันธ์เชิงพหุเป็นส่วนขยายของการวิเคราะห์การถดถอยและสหสัมพันธ์อย่างง่าย นั่นคือแทนที่จะพิจารณาปัจจัยที่เกี่ยวข้องเพียงปัจจัยเดียวในการวิเคราะห์เกี่ยวกับเรื่องที่ต้องการศึกษา ก็จะพิจารณาปัจจัยที่เกี่ยวข้องตั้งแต่ 2 ปัจจัยขึ้นไป

3.1 เงื่อนไขเบื้องต้นในการวิเคราะห์การถดถอยเชิงพหุ

1.ตัวแปรอิสระและตัวแปรตามมีความสัมพันธ์ในแบบเชิงเส้น สมการเส้นถดถอยแบบเส้นตรงที่มีตัวแปรอิสระ k ตัวจะอยู่ในรูป

$$Y_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + ... + \beta_k X_{ki} + e_i$$

โดยที่ Y_i = ค่าของตัวแปรตามที่ i ที่เก็บรวบรวมได้

α = ค่าคงที่ในสมการถดถอย ซึ่งเป็นค่าของ y เมื่อทุกค่าของ $X_k = 0$

$\beta_1, \beta_2, ..., \beta_k$ = สัมประสิทธิ์ของเส้นถดถอยซึ่งสอดคล้องกับแต่ละค่าของตัวแปรอิสระ

e_i = ความคลาดเคลื่อนสุ่ม ซึ่งเกิดจากการสุ่มตัวอย่าง

2. ความแปรปรวนของตัวแปรตามเมื่อกำหนดค่าต่าง ๆ ของตัวแปรอิสระจะมีค่าเท่ากัน เงื่อนไขนี้เรียกว่า มีคุณสมบัติของ homoscedasticity

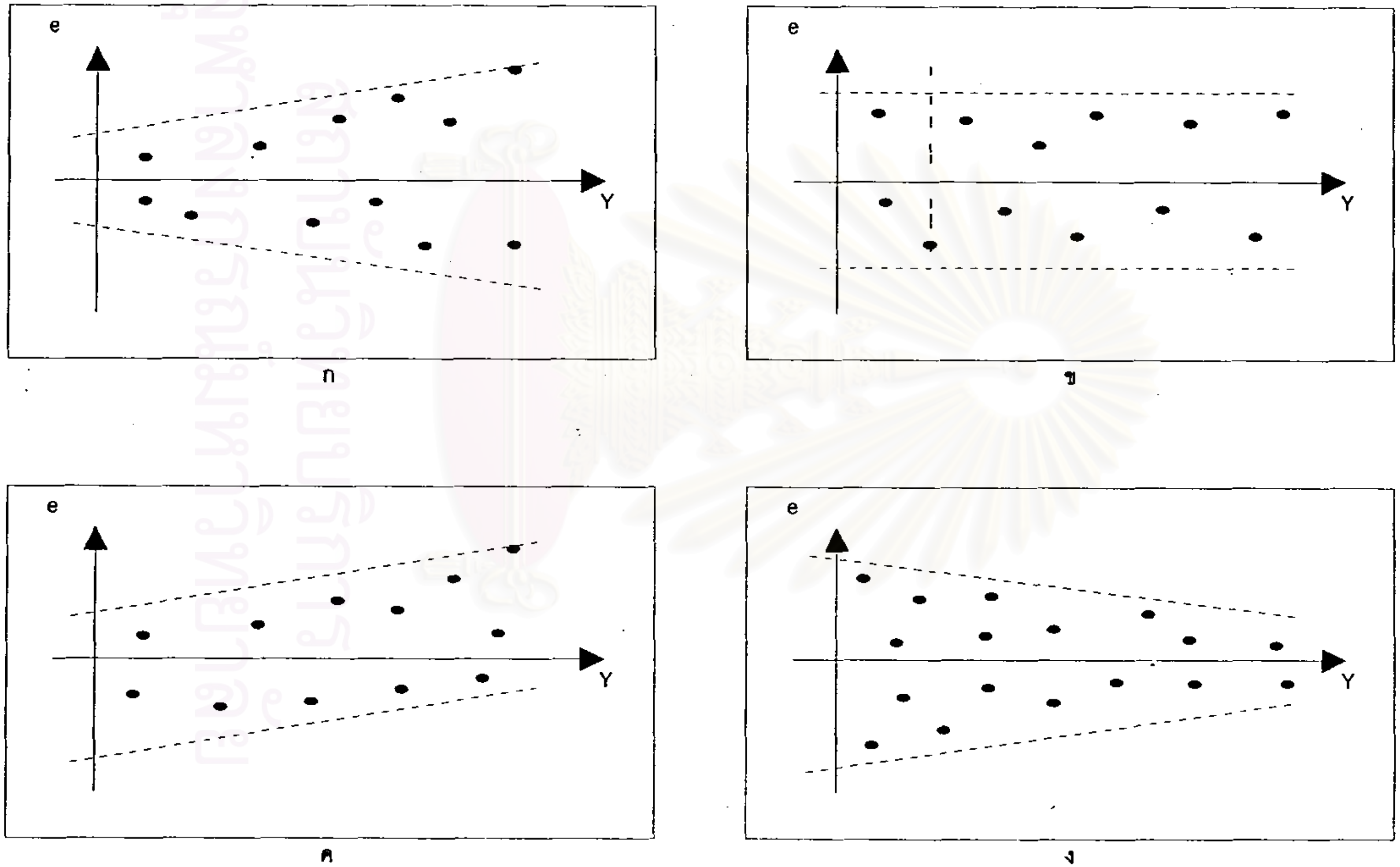
e เป็นความคลาดเคลื่อนที่มีค่าเฉลี่ยเป็นศูนย์ E(e) = 0 และ V(e) = σ^2

การตรวจสอบว่า E(e) = 0

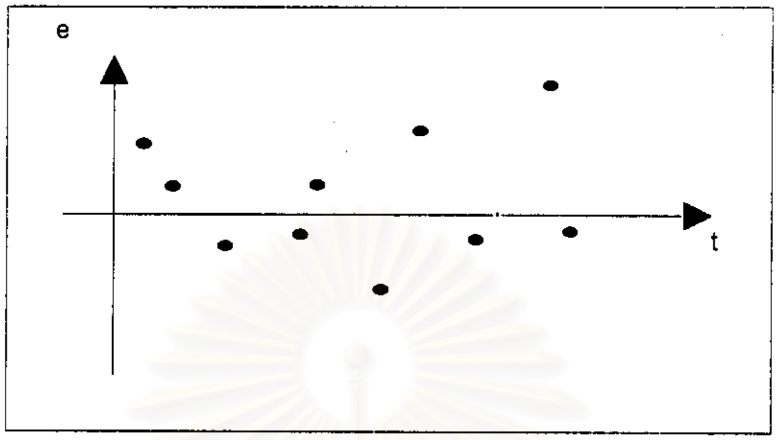
เนื่องจากเราใช้วิธีกำลังสองน้อยที่สุดในการประมาณ β_0 ด้วย a และ β_i ด้วย b_i ซึ่งจะทำให้ $\sum e_i = 0$

การตรวจสอบว่า V(e) = σ^2 การตรวจสอบว่า V(e) = V(Y) = σ^2 = ค่าคงที่ จะทำโดยการเขียนกราฟแสดงความสัมพันธ์ระหว่าง e กับ $\bar{Y}$ ถ้า V(e) ไม่เท่ากับค่าคงที่จะเรียกว่าเกิดปัญหา Heteroscedatic ดังแสดงในรูปที่ 3-1ก แต่ถ้า V(e) = ค่าคงที่ตามสมมติฐานที่ตั้งไว้ จะเรียกว่า Homoscedatic ดังแสดงในรูปที่ 3-1ข และ 3-1ค และ 3-1ง

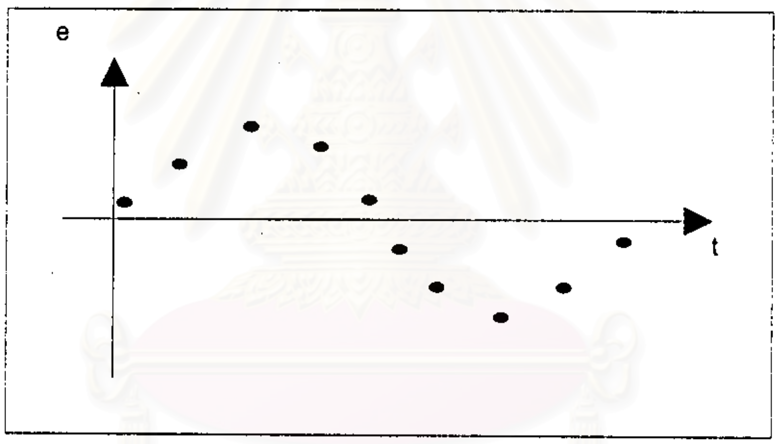
กราฟ 3-1 การตรวจสอบความแปรปรวน



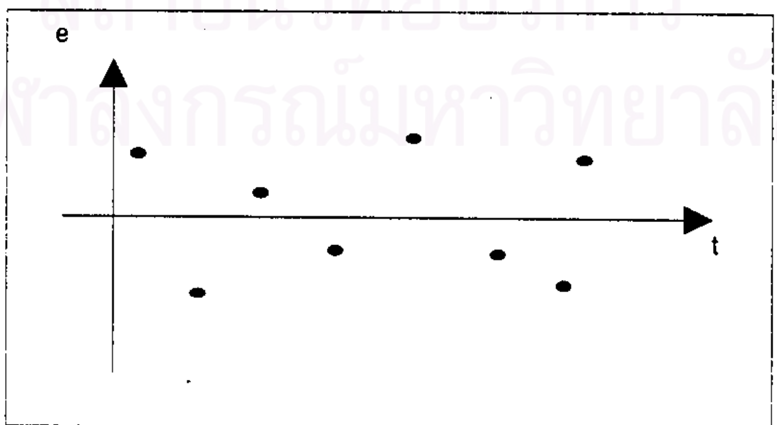
กราฟ 3-2 การตรวจสอบ ei และej เป็นอิสระกันโดยการเขียนกราฟ



ก



ข



ค

จากรูปที่ 3-2 (ก) จะพบว่า σ_e^2 จะมีค่าน้อยเมื่อ $\bar{Y}$ มีค่าน้อย และเมื่อ $\bar{Y}$ มีค่ามาก σ_e^2 จะมีค่ามากด้วย ในขณะที่รูปที่ 3-2 (ข) (ค) ค่า σ_e^2 จะคงที่เมื่อ Y เปลี่ยนไป ส่วนรูปที่ 3-2 (ง) ค่า σ_e^2 จะมีค่าน้อยเมื่อ Y มีค่ามาก

3.ค่าของตัวแปรสุ่มตัวที่อยู่ติดกันไม่สัมพันธ์กัน ถ้าค่าของตัวแปรสุ่มไม่สอดคล้องกับคุณสมบัติข้อนี้ เรียกว่า ข้อมูลมี autocorrelation กัน ข้อมูลอนุกรมเวลาหรือข้อมูลที่เก็บรวบรวมตามลำดับเวลาที่เกิดขึ้นมักจะมี autocorrelation กัน สำหรับการวิเคราะห์การถดถอยเชิงพหุ autocorrelation จะมีผลต่อความแม่น (precision) ของการทดสอบสมมติฐานและการประมาณค่าในแบบช่วง แต่จะไม่มีผลสำหรับความแม่นในการประมาณค่าแบบจุด

การตรวจสอบ e_i และ e_j เป็นอิสระกัน การตรวจสอบความเป็นอิสระกันของ e_i และ e_j โดยที่ $e_i = Y_i - Y_i$ และ $e_j = Y_j - Y_j$ อาจทำได้ 2 วิธี คือ

- โดยการเขียนกราฟแสดงความสัมพันธ์ระหว่าง e_i กับ t ถ้า e_i และ e_j มีความสัมพันธ์จะเรียกว่าเกิด Autocorrelation ดังแสดงในรูปที่ 3-2 (ข) และ (ค) โดยรูปที่ 3-2 (ข) แสดงความสัมพันธ์ระหว่าง e_t และ e_{t+j} ในทางบวกซึ่งเรียกว่าเกิด Positive Autocorrelation ส่วนรูปที่ 3-2 (ค) แสดงความสัมพันธ์ในทิศทางตรงข้ามหรือทางลบซึ่งเรียกว่า Negative Autocorrelation ส่วนรูปที่ 3-2 (ก) แสดงว่าค่าคลาดเคลื่อนเป็นอิสระกัน

-ใช้สถิติทดสอบ Durbin - Watson การทดสอบความอิสระกันของค่าคลาดเคลื่อนเมื่อใช้การทดสอบของ durbin - Watson เป็นการทดสอบความสัมพันธ์ของ e_t และ e_{t-1} โดยที่ t เป็นช่วงเวลา

$$\text{สถิติทดสอบ Durbin - Watson} = d = \frac{\sum_{2}^{n}(e_t - e_{t-1})^2}{\sum_{1}^{n} e_t^2}$$

โดยที่ $0 \leq d \leq 4$ และมีคุณสมบัติดังนี้

-ถ้าค่าความคลาดเคลื่อน (e_t) แต่ละตัว เป็นอิสระกัน ค่า d จะมีค่าใกล้ 2

-ถ้า $d < 2$ จะแสดงถึงความสัมพันธ์ในทางบวกของค่าคลาดเคลื่อน และถ้า $d \approx 0$ ความสัมพันธ์ของ e_t แต่ละตัวจะมาก

-ถ้า $d > 0$ จะแสดงถึงความสัมพันธ์ของค่าคลาดเคลื่อนในทางลบ และถ้า $d \approx 4$ ความสัมพันธ์ของ e_t แต่ละตัวจะมาก

4.การแจกแจงของตัวแปรตาม Y เป็นการแจกแจงแบบปกติ (Normal) นั่นคือการแจกแจงของ e_t ในการถดถอยเชิงพหุจะเป็นการแจกแจงแบบปกติด้วย ซึ่งการทดสอบว่า e_t มี

การแจกแจงแบบปกติหรือไม่ ทำได้ดังนี้ ใช้ Chi-Square Test ใช้ Kolmogorov - Smirnov Test หรือ การตรวจสอบด้วยการวิเคราะห์ค่า Residual (Residual Analysis)

$$\text{Residual} = \frac{\text{Error}}{\sigma}$$

$$\sigma^2 = \frac{SSE}{n-p}$$

โดยที่ ถ้าค่า Residual อยู่ระหว่าง -2 ถึง +2 จะสามารถสรุปได้ว่า e_i มีการแจกแจงแบบปกติ

3.2 รูปแบบของสมการความถดถอยเชิงซ้อน

ถ้ามีตัวแปรอิสระ k ตัว ($X_1, X_2, \ldots, X_k$) ที่มีความสัมพันธ์กับตัวแปรตาม Y โดยที่ความสัมพันธ์อยู่ในรูปเชิงเส้น จะได้สมการความถดถอยเชิงซ้อน ซึ่งแสดงความสัมพันธ์ระหว่าง Y และ $X_1, X_2, \ldots, X_k$ ดังนี้

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + e$$

โดยที่ β_0 = ส่วนตัดแกน Y เมื่อกำหนดให้ $X_1 = X_2 = \ldots = X_k = 0$

$\beta_1, \beta_2, \ldots, \beta_k$ เป็นสัมประสิทธิ์ความถดถอยเชิงส่วน (Partial Regression Coefficient) โดยที่ β_i เป็นค่าที่แสดงถึงการเปลี่ยนแปลงของตัวแปรตาม Y เมื่อตัวแปรอิสระ X_i เปลี่ยนไป 1 หน่วย โดยที่ตัวแปรอิสระ X ตัวอื่น ๆ มีค่าคงที่

3.3 การทดสอบสมการความถดถอยเชิงซ้อนโดยใช้การวิเคราะห์ความแปรปรวนจำแนกแบบทางเดียว

จากสมการความถดถอยเชิงซ้อน

$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + e$

ค่าความแปรปรวนของ Y = ค่าแปรปรวนที่เกิดจากอิทธิพลของ $X_1, X_2, \ldots, X_k$ +e

หรือ SST = SSR + SSE

โดยที่ SST (Sum Square of Total) คือ ค่าแปรปรวนทั้งหมดของ Y

หรือ $SST = \sum_{i=1}^{n} (Y_i - \bar{y})^2$

SSR (Sum Square of Regression) คือ ค่าแปรปรวนของ Y เนื่องจากอิทธิพลของ $X_1, \ldots, X_k$

SSE (Sum Square of Error) คือ ค่าแปรปรวนของ Y เนื่องจากอิทธิพลอื่น ๆ หรือเรียกว่า ค่าแปรปรวนอย่างสุ่ม

ตารางที่ 3-1 การวิเคราะห์ความแปรปรวนของการวิเคราะห์ความถดถอยเชิงซ้อน

SV	DF	SS	MS	F
ความถดถอย (Regression)	k	SSR	MSR = SSR/k	$\frac{MSR}{MSE}$
ความคลาดเคลื่อน (Error)	n-k-1	SSE	MSE = SSE/(n-k-1)	
ผลรวม (Total)	n-1	SST		

โดยที่ $SSR = \underline{b}'X'\underline{Y} - n\bar{y}^2$

$$SST = \sum_{i=1}^{n}(Y_i - \bar{y})^2 = \underline{Y'Y} - n\bar{y}^2$$

$$SSE = \sum_{i=1}^{n}\left[Y_i - (a + b_1X_1 + b_2X_2 + \ldots + b_kX_k)\right]^2$$

หรือ $SSE = SST = \underline{Y}'\underline{Y} - \underline{b}'X'\underline{Y}$

3.4 การประมาณค่าความคลาดเคลื่อนของการถดถอย (Estimation of Standard Deviation of Regression)

ในกรณีที่มีตัวแปรอิสระ k ตัว จะได้ค่าแปรปรวนของการประมาณ คือ

$$S_e^2 = S_{Y.12\ldots k}^2 = S^2$$

โดยที่ $$S^2 = \frac{SSE}{n-k-1} = \frac{\sum_{i=1}^{n}(Y_i - \bar{Y}_i)^2}{n-k-1}$$

ดังนั้น ความคลาดเคลื่อนหรือค่าเบี่ยงเบนมาตรฐานของค่าประมาณ คือ

$$S = \sqrt{S^2} = \sqrt{\frac{SSE}{(n-k-1)}} = \sqrt{MSE}$$

ในกรณีที่มีตัวแปรอิสระ 2 ตัว คือ X_1 และ X_2 (k =2) จะได้

$$S^2 = \frac{\sum_{i=1}^{n}\left(Y_i - \overline{Y}_i\right)^2}{n-2-1} = MSE$$

3.5 การประมาณค่าเบี่ยงเบนมาตรฐานของ b_i

ค่าเบี่ยงเบนมาตรฐานของ b_1 คือ

$$S_{b1} = \sqrt{S^2_{b_1}} \qquad = \frac{S}{\sqrt{\sum x_1^2\left(1-r_{12}^2\right)}}$$

$$S_{b2} = \sqrt{S^2_{b_2}} \qquad = \frac{S}{\sqrt{\sum x_1^2\left(1-r_{12}^2\right)}}$$

โดยที่ $x_i = \left(X_{ij} - \overline{X}_i\right)$

r_{12} = สัมประสิทธิ์สหสัมพันธ์อย่างง่ายระหว่าง X_1 และ X_2

และ $r_{12} = \frac{\sum x_1 x_2}{\sqrt{\left(\sum x_1^2\right)\left(\sum x_2^2\right)}}$

3.6 การทดสอบสมมติฐานเกี่ยวกับสัมประสิทธิ์ความถดถอย

ถ้ามีตัวแปรอิสระ k ตัว (x_1 , x_2 ...., x_k) ที่มีความสัมพันธ์กับตัวแปรตาม Y และเมื่อได้ทดสอบ F-test จากตารางการวิเคราะห์ความแปรปรวนเพื่อทดสอบสมมติฐาน

$H_0 : \beta_1 = \beta_2 = ...\beta_k = 0$

H_0 : มี β_i อย่างน้อย 1 ค่าที่ $\neq 0$; i = 1, 2,..., k

ผลของการทดสอบสมมติฐานข้างต้น โดยการใช้สถิติ F จะเป็น

-ยอมรับสมมติฐาน H_0 ถ้า $F < F_{k,\,n-k-1}$ แสดงว่าตัวแปร Y ไม่มีความสัมพันธ์กับตัวแปรอิสระทั้ง k ตัว (X_1 , X_2 ,...., X_k)

-ปฏิเสธสมมติฐาน H_0 ถ้า $F > F_{k,\,n-k-1}$ แสดงว่าตัวแปรอิสระ (X's) อย่างน้อย 1 ตัวที่มีความสัมพันธ์กับ Y

ในกรณีที่เกิด ข. คือ ปฏิเสธ H_0 จะต้องทำการทดสอบต่อไปว่า β_i ตัวใดบ้างที่ไม่เท่ากับศูนย์ หรือมี X ตัวใดบ้างที่สัมพันธ์กับ Y โดยการทดสอบสมมติฐานดังต่อไปนี้

สมมติฐาน $H_0 : \beta_i = 0$ $H_0 : \beta_i \neq 0 ; i = 1, 2,\ldots, k$

สถิติทดสอบ $t = \dfrac{b_i - 0}{S_{b_i}}$ หรือใช้สถิติทดสอบ Z ถ้า n มีค่ามาก

ผลของการทดสอบสมมติฐานข้างต้น

- เขตปฏิเสธสมมติฐาน H_0 จะปฏิเสธสมมติฐาน H_0 เมื่อ $t > t_{1-\alpha/2;n-k-1}$

หรือ กล่าวว่าจะปฏิเสธ H_0 ถ้า $|t| > t_{1-\alpha/2;n-k-1}$

ตาราง 3-2 รูปแบบการทดสอบสมมติฐานของสัมประสิทธิ์การถดถอย (β)

รูปการทดสอบสมมติฐานเกี่ยวกับสัมประสิทธิ์ความถดถอย $\beta_i ; i = 1, 2,\ldots, k$	
การทดสอบข้างเดียว	การทดสอบแบบสองข้าง
$H_0 : \beta_i = 0$	$H_0 : \beta_i = 0$
$H_1 : \beta_i < 0$	$H_1 : \beta_i \neq 0$
(หรือ $H_1 : \beta_i > 0$)	
สถิติทดสอบ	สถิติทดสอบ
$t = b_i / S_{bi}$	$t = b_i / S_{bi}$
เขตปฏิเสธสมมติฐาน H_0 $t < t_{1-\alpha;n-k-1}$	เขตปฏิเสธสมมติฐาน H_0 $t > t_{1-\alpha/2;n-k-1}$
(หรือ $t > t_{1-\alpha;n-k-1}$ เมื่อ $H_1 : \beta_i > 0$)	หรือ $t < t_{1-\alpha/2;n-k-1}$ หรือ $\lvert t\rvert = t_{1-\alpha/2;n-k-1}$

3.7 สัมประสิทธิ์การตัดสินใจเชิงซ้อน (Multiple Coefficient of Determination : R^2 หรือ r^2)

สัมประสิทธิ์การตัดสินใจเชิงซ้อนจะมีความหมายเหมือนกับความหมายของสัมประสิทธิ์การตัดสินใจ คือเป็นสัดส่วนหรือเปอร์เซ็นต์ที่ตัวแปรอิสระ ($X_1, X_2, \ldots, X_k$) สามารถอธิบายการเปลี่ยนแปลงของ Y ได้ หรือกล่าวได้ว่าสัมประสิทธิ์การตัดสินใจเชิงซ้อนเป็นสัดส่วนหรือเปอร์เซ็นต์ของความผันแปร Y ที่มีสาเหตุเนื่องจากความผันแปรของ $X_1, X_2, \ldots$ และ X_k โดยที่สัมประสิทธิ์การตัดสินใจเชิงซ้อนจะให้สัญญลักษณ์ $R^2_{Y.123\ldots k}$ แต่โดยทั่วไปจะใช้ R

$$r^2 = R^2 = \frac{\text{ความผันแปรของ Y เนื่องจากอิทธิพลของ } X_1, X_2, \ldots, X_k}{\text{ความผันแปรทั้งหมด}}$$

$$= SSR / SST$$

หรือ $r^2 = R^2 = (SST - SSE) / SST = 1 - SSE / SST$

โดยที่ $0 \le R^2 \ , \ r^2 \le 1$

ถ้าค่า R^2 ที่ใกล้ 1 จะหมายถึง $X_1 , X_2,..., X_k$ มีความสัมพันธ์กับ Y มาก แต่ถ้า R^2 เข้าใกล้ศูนย์ หมายถึง ค่า $X_1 , X_2,..., X_k$ มีความสัมพันธ์กับ Y น้อย

เนื่องจาก SSR จะเพิ่มขึ้นถ้าเพิ่มตัวแปรอิสระ เช่น เดิมมี X_1 และ X_2 ที่มีความสัมพันธ์กับ Y แต่ถ้าเพิ่มตัวแปรอิสระ X_3 เข้าในสมการความถดถอย จะได้ว่า

$$SSR(X_1, X_2, X_3) > SSR(X_1 , X_2)$$

โดยที่ $SSR(X_1, X_2, X_3)$ หมายถึง SSR ของสมการความถดถอยที่มีตัวแปรอิสระ X_1, X_2 และX_3

$$Y = \beta_0+\beta_1x_1+\beta_2x_2+\beta_3x_3+e$$

และ $SSR(X_1 , X_2)$ หมายถึง SSR ของสมการถดถอยที่มีตัวแปรอิสระ X_1 และ X_2

$$Y = \beta_0+\beta_1x_1+\beta_2x_2+e$$

ดังนั้นเมื่อเพิ่มตัวแปรอิสระเข้าสมการความถดถอยจะทำให้ค่า R^2 มากขึ้นทั้งที่ตัวแปรอิสระ X ที่เพิ่มอาจจะไม่มีความสัมพันธ์กับ Y เลยก็ได้ จึงมีการปรับค่า R^2 ให้ถูกต้องขึ้น เรียกว่า Adjusted R^2 โดยที่

$$R_a^2 = \text{Adjusted } R^2$$

$$R_a^2 = 1 - \frac{SSE/(n-k-1)}{SST/(n-1)}$$

หรือ $$R_a^2 = 1 + \frac{(n-1)}{n-k-1}\left(R^2 - 1\right)$$

ค่าสัมประสิทธิ์สหพันธ์เชิงซ้อน ได้จากการถอดรากที่สองของสัมประสิทธิ์การตัดสินใจเชิงซ้อน สัมประสิทธิ์สหสัมพันธ์เชิงซ้อน = $R_{Y.12...k} = R = \sqrt{R^2_{Y.123.k}}$ โดยที่ $0 \leq R \leq 1$ โดยที่สัมประสิทธิ์สหสัมพันธ์เชิงซ้อนแสดงถึงความสัมพันธ์ระหว่าง Y กับ $X_1, X_2, ..., X_k$ ดังนี้

1.R มีค่าเข้าใกล้ศูนย์ แสดงว่า Y มีความสัมพันธ์กับ $X_1, X_2, ..., X_k$ น้อยมาก และถ้า R = 0 แสดงว่า Y ไม่มีความสัมพันธ์กับ $X_1, X_2, ..., X_k$ เลย

2.R มีค่าเข้าใกล้ 1 แสดงว่า Y มีความสัมพันธ์กับตัวแปรอิสระทั้ง k ตัวมีมาก

3.8 การปรับสมการความถดถอยเมื่อความคลาดเคลื่อนไม่เป็นไปตามข้อกำหนด

เมื่อความคลาดเคลื่อน e ไม่ได้มีการแจกแจงแบบปกติ หรือ ค่าแปรปรวนของ e ไม่ใช่ค่าคงที่ หมายถึงสมการความถดถอยเชิงเส้นนั้นไม่เหมาะสม การแก้ไขทำได้โดยการเปลี่ยนแปลงเมื่อรูปแบบของตัวแปรตาม Y ดังนี้

1. เมื่อ V(e) เพิ่มขึ้นเมื่อ Y เพิ่มขึ้น หรือเมื่อ การแจกแจงของความคลาดเคลื่อน e ไม่ได้มีการแจกแจงแบบปกติ แต่เบ้ซ้าย ในกรณีนี้ จะให้ $Y' = \log(Y)$ โดยที่ $Y > 0$ และตัวแปรตามคือ Y' นั่นคือ $Y' = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k + e$

2. เมื่อ V(e) เป็นสัดส่วนกับ E(Y) หรือ e มีการแจกแจงเบ้ขวา จะต้องให้ $Y' = Y^2$ สมการความถดถอย คือ $Y' = Y^2 = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k + e$

3.9 การเกิดปัญหา Multicollinearity

การวิเคราะห์ความถดถอยเชิงซ้อนซึ่งศึกษาถึงความสัมพันธ์ระหว่างตัวแปรตาม (Y) กับตัวแปรอิสระหลายตัว ($X_1, X_2, ..., X_k$; $k \geq 2$) นั้น มีข้อกำหนดว่า ตัวแปรอิสระเหล่านั้นจะต้องไม่มีความสัมพันธ์กัน แต่ในทางปฏิบัติพบว่า ตัวแปรอิสระมักจะมีความสัมพันธ์กันเอง เช่น Y = ยอดขาย X_1 = ราคาขายต่อหน่วย X_2 = จำนวนคู่แข่ง X_3 = ค่าโฆษณา และ X_4 = ค่าโฆษณาของคู่แข่ง จะพบว่า X_i มีความสัมพันธ์กัน การที่ตัวแปร X มีความสัมพันธ์กันจะทำให้เกิดปัญหาที่เรียกว่า Multicollinearity การเกิดปัญหา Multicollinearity จะมากหรือน้อยจะขึ้นอยู่กับความสัมพันธ์ระหว่างตัวแปรอิสระ X ถ้าตัวแปรอิสระมีความสัมพันธ์กันมาก ปัญหา Multicollinearity จะมากด้วย ซึ่งทำให้ผลของการเกิดปัญหา Multicollinearity รุนแรงด้วย

<u>ผลของการเกิดปัญหา Multicollinearity</u>

การที่ตัวแปรอิสระ X มีความสัมพันธ์กันจะทำให้ผลลัพธ์ของการวิเคราะห์ความถดถอยผิดไป ดังนี้

1. ทำให้ค่าเบี่ยงเบนมาตรฐานของสัมประสิทธิ์ความถดถอย S_b มีค่าสูงมาก ซึ่งมีผลทำให้สถิติทดสอบ t มีค่าต่ำกว่าปกติ โดยที่ในการทดสอบสัมประสิทธิ์ความถดถอย

$H_0 : \beta_i = 0$; $i = 1, 2, ..., k$

$H_1 : \beta_i \neq 0$

สถิติการทดสอบ $t = \frac{b_i}{S_{b_i}}$

ซึ่งเป็นการทดสอบว่า X_i มีความสัมพันธ์กับ Y หรือไม่ เมื่อสถิติทดสอบ t มีค่าต่ำกว่าปกติ เนื่องจาก S_b มีค่าสูงมากจึงทำให้ยอมรับสมมติฐาน H_0 หรือสรุปว่า X_i มีความสัมพันธ์กับ Y อย่างมาก นอกจากนั้นเมื่อใช้สถิติทดสอบ F ในการทดสอบสมมติฐาน จะได้ผลสรุปตรงข้ามกับสถิติทดสอบ t

$H_0 : \beta_1 = \beta_2 = 0$

H_1 : มี β_i อย่างน้อย 1 ค่า $\neq 0$

ผลของการทดสอบจะเป็นปฏิเสธ H_0 นั่นคือ β_i อย่างน้อย 1 ค่าไม่เท่ากับศูนย์ หรือมี X อย่างน้อย 1 ตัวที่มีความสัมพันธ์กับ Y แต่เมื่อทำการทดสอบ β ทีละตัวหรือทดสอบความสัมพันธ์ระหว่าง Y กับ X ทีละตัวโดยใช้สถิติทดสอบ t จะยอมรับสมมติฐาน $H_0 : \beta_i = 0$ ทุกค่าของ i ซึ่งไม่มี X_i ตัวใด ซึ่งมีความสัมพันธ์กับ Y

2.การที่ตัวแปรอิสระ X มีความสัมพันธ์กันจะทำให้เครื่องหมายของสัมประสิทธิ์ความถดถอย (β, b) ตรงข้ามกับที่ควรจะเป็น

3.การที่ตัวแปรอิสระมีความสัมพันธ์กัน จะทำให้ค่าสัมประสิทธิ์ความถดถอย (β, b) เปลี่ยนแปลงไป (ไม่คงที่) เมื่อมีตัวแปรอิสระเพิ่มขึ้น

เมื่อตัวแปรอิสระ X ไม่มีความสัมพันธ์กัน สัมประสิทธิ์ความถดถอยจะไม่เปลี่ยนแปลงไปเมื่อมีตัวแปร X เพิ่มขึ้น

3.10 การป้องกันการเกิดปัญหา Multicollinearity

เมื่อเกิดปัญหา Multicollinearity จะทำให้สมการความถดถอยไม่ถูกต้อง ดังนั้นผู้วิเคราะห์จึงต้องหาทางป้องกันการเกิดปัญหาดังกล่าว สำหรับวิธีการป้องกันอาจทำได้ดังนี้

1.คำนวณหาสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรอิสระ (X) ต่าง ๆ แล้วทำการทดสอบสมมติฐานว่า สัมประสิทธิ์สหสัมพันธ์ ρ ของ X แต่ละคู่เป็นศูนย์หรือไม่ ถ้าผลการทดสอบยอมรับว่า ρ ของแต่ละคู่เป็นศูนย์ แสดงว่าตัวแปรอิสระ (X) ต่าง ๆ ไม่มีความสัมพันธ์กัน จะพบว่าในทางปฏิบัตินั้นการที่จะหาตัวแปร X ที่เป็นอิสระกันทุกคู่เป็นไปได้ยากเนื่องจากตัวแปร X มักจะมีความสัมพันธ์กัน

2.ใช้วิธี Stepwise ซึ่งเป็นวิธีการเลือกตัวแปรอิสระเข้าสมการความถดถอย โดยวิธี Stepwise มีหลักเกณฑ์ว่า จะนำตัวแปรอิสระเข้าสมการความถดถอยครั้งละ 1 ตัว ถ้าตัวแปรอิสระ

ที่นำเข้ามีความสัมพันธ์กับตัวแปรอิสระที่มีอยู่แล้วในสมการถดถอย วิธี Stepwise จะตัดตัวแปรอิสระที่สัมพันธ์กันตัวใดตัวหนึ่งออกจากสมการความถดถอย

3.11 การเลือกตัวแปรอิสระเข้าสมการความถดถอย

เนื่องจากการวิเคราะห์ความถดถอยเป็นการศึกษาถึงความสัมพันธ์ระหว่างตัวแปรตามกับตัวแปรอิสระ เพื่อที่จะพยากรณ์หรือประมาณค่าตัวแปรตามโดยที่ทราบค่าของตัวแปรอิสระ ค่าประมาณจะมีความถูกต้องหรือไม่ จึงขึ้นอยู่กับตัวแปรอิสระที่เลือกเข้าสมการ สมการความถดถอยที่ดีคือสมการความถดถอยที่ประกอบด้วยตัวแปรอิสระที่เหมาะสมซึ่งจะให้ค่าพยากรณ์ของตัวแปรตาม Y ใกล้เคียงกับค่าจริง การพิจารณาว่าตัวแปรอิสระ X ตัวใด มีความสัมพันธ์กับ Y นั้น จะพิจารณาจาก สถิติทดสอบ F สถิติทดสอบ t สัมประสิทธิ์การตัดสินใจ R^2 สัมประสิทธิ์สหสัมพันธ์ r และความคลาดเคลื่อนมาตรฐาน ($S_{y,x}$)

การเลือกสมการความถดถอยที่ดีที่สุด (Selecting the Best Regression Equation) สามารถทำได้ 2 ขั้นตอน ดังนี้

1.การเลือกรูปแบบสมการความถดถอย

ผู้วิเคราะห์ต้องเลือกรูปแบบที่แสดงความสัมพันธ์ระหว่าง Y กับ $X_1, X_2, ..., X_k$ ที่ถูกต้องโดยพิจารณาจากค่าความเคลื่อนว่าเป็นไปตามข้อกำหนดหรือไม่ ซึ่งรูปแบบอาจจะไม่ได้อยู่ในรูปเชิงเส้น แต่อยู่ในรูปกำลังสองหรือ ln ฯลฯ เช่น ถ้า Y กับ X_1 และ X_2 มีความสัมพันธ์ในรูปกำลังสอง

$$Y = \beta_0 + \beta_1 x_1^2 + \beta_2 x_2^2 + \beta_3 x_1 x_2 + e$$

2.การเลือกตัวแปรอิสระที่เหมาะสมกับสมการความถดถอย

การเลือกตัวแปรอิสระ X ที่มีความสัมพันธ์กับ Y เพื่อให้ได้สมการความถดถอยที่ดีที่สุด มีหลายวิธี ดังนี้

เทคนิคการเลือกตัวแปรอิสระที่เหมาะสมสำหรับสมการความถดถอย การเลือกตัวแปรอิสระ X เข้าสมการความถดถอยนั้น ก่อนอื่นผู้วิเคราะห์จะต้องพิจารณาว่ามีตัวแปรอิสระใดบ้างที่น่าจะมีความสัมพันธ์กับ Y ทั้งสัมพันธ์ในทางบวกและลบ

1.All Possible Regression

2.Backward Elimination

3.Forward Selection

4.Stepwise Regression วิธีการคัดเลือกตัวแปรอิสระ โดยวิธี Stepwise Regression นี้ เป็นการคัดเลือกตัวแปรอิสระที่ค่อนข้างจะซับซ้อน ดังนี้ คือ ขั้นแรกจะเป็นการคัด

เลือกตัวแปรอิสระเข้าไปในสมการทีละตัวแปร โดยพิจารณาเลือกตัวแปรอิสระที่มีความสัมพันธ์สูงสุดกับตัวแปรตามเข้าไปในสมการ แล้วจึงพิจารณาตัวแปรอิสระที่มีความสัมพันธ์เชิงส่วน (ความสัมพันธ์ที่กำหนดให้ตัวแปรอิสระที่เข้าไปในสมการแล้วคงที่) กับตัวแปรตามมากที่สุดเข้าไปในสมการ และพร้อมกันนั้นก็พิจารณาว่าตัวแปรอิสระที่เข้าไปในสมการก่อนหน้านั้นทุกตัวแปรยังควรอยู่ในสมการอีกหรือไม่ ถ้าไม่ควรอยู่ก็ตัดออกและดำเนินการคัดเลือกตัวแปรอิสระใหม่ ถ้าควรอยู่ก็ดำเนินการกัดเลือกตัวแปรอิสระใหม่ วิธี Stepwise Regression จะมีการพิจารณาตัวแปรที่ถูกคัดเลือกเข้าไปอยู่ในสมการใหม่อีกทุกครั้งที่มีการเพิ่มตัวแปรอิสระเข้าไปใหม่ เพราะวิธี Stepwise Regression ถือว่าเมื่อตัวแปรอิสระนั้นอาจจะมีผลต่อสมการแตกต่างจากเดิมก็ได้ การคัดเลือกตัวแปรอิสระจะดำเนินการไปจนกระทั่งไม่มีตัวแปรอิสระใด เข้าไปหรือถูกตัดออกจากสมการได้อีก ซึ่งวิธีการคัดเลือกตัวแปรอิสระโดยวิธีอิสระใด เข้าไปหรือถูกตัดออกจากสมการได้อีก ซึ่งวิธีการคัดเลือกตัวแปรอิสระโดยวิธี Stepwise Regression นั้น อาจกำหนดเป็นขั้นตอนได้ดังนี้

ขั้นที่ 1 คำนวณค่าสัมประสิทธิ์สหสัมพันธ์ r ระหว่างตัวแปรตามกับตัวแปรอิสระแต่ละตัว

ขั้นที่ 2 เลือกตัวแปรอิสระที่มีค่า r สูงสุด แล้วจึงกำหนดรูปแบบของสมการดังนี้

$Y = \beta_0 + \beta_p x_p + \varepsilon$ (X_p คือ ตัวแปรอิสระที่ถูกเลือกเข้าไป)

ขั้นที่ 3 ทดสอบสมมติฐาน $H_0 = \beta_p = 0$ โดยค่าสถิติ F - test ถ้ายอมรับสมมติฐานให้หยุดและถือว่าไม่มีตัวแปรอิสระใดถูกเลือก ถ้าปฏิเสธสมมติฐานให้ดำเนินการคัดเลือกตัวแปรอิสระต่อไป

ขั้นที่ 4 คำนวณค่าสัมประสิทธิ์เชิงส่วน (Partial Correlation Coefficient) $r_{yi.123...k}$ ระหว่างตัวแปรตามกับตัวแปรอิสระที่เหลือ แต่ละตัว โดยกำหนดให้ตัวแปรอิสระที่อยู่ในสมการมีค่าคงที่

จากเทคนิคทั้ง 4 วิธี Stepwise เป็นวิธีที่ป้องกันการเกิดปัญหา Multicollinearity ได้ เนื่องจากเมื่อมีตัวแปรอิสระที่มีความสัมพันธ์กันอยู่ในสมการความถดถอย จะทำให้สัมประสิทธิ์ความถดถอยเปลี่ยนไป ค่าสถิติทดสอบ t เปลี่ยนไป ซึ่งอาจมีผลให้ตัดตัวแปรอิสระที่สัมพันธ์กันบางตัวออกไป สำหรับวิธี All Possible Regression นั้นผู้วิเคราะห์จะต้องเป็นผู้ตัดสินใจเองว่าควรจะเลือกตัวแปรอิสระตัวใดบ้าง ส่วนวิธี Backward และวิธี Forward ก็เป็นวิธีที่มีหลักเกณฑ์ที่ดีระดับหนึ่ง