บทที่ 1
บทนำ

ความเป็นมาและความสำคัญของปัญหา

ในสังคมปัจจุบัน เหตุการณ์ต่าง ๆ ได้เปลี่ยนไปมากมาย จนไม่อาจคาดคะเนเหตุการณ์ที่เกิดขึ้นได้อย่างถูกต้องว่ามีสาเหตุมาจากอะไร ฉะนั้นสถาบันต่าง ๆ จึงมุ่งเน้นให้ผู้เรียนรู้จักวิธีการทำวิจัย เพื่อการค้นหาคำตอบในด้านการพยากรณ์มากขึ้น โดยเฉพาะการวิจัยทางการศึกษาและจิตวิทยา ซึ่งจะเห็นได้จากผลงานวิจัยที่ปรากฏ ทั้งนี้เพราะการพยากรณ์มีบทบาทสำคัญมากทั้งในด้านการวางแผนและการตัดสินใจ

วิธีการพยากรณ์ที่นิยมใช้กันมากในปัจจุบัน คือ การวิเคราะห์การถดถอย (Regression Analysis) เพราะนอกจากจะเป็นการพยากรณ์ในเชิงสถิติที่มีมาตรการวัดความถูกต้องและทดสอบนัยสำคัญของตัวแบบสมการและของค่าพยากรณ์แล้ว ยังสนองหลักความเชื่อที่ว่าสิ่งที่จะพยากรณ์ถูกกำหนดให้เป็นไปโดยปัจจัยต่าง ๆ ที่สัมพันธ์กันอีกด้วย การวิเคราะห์การถดถอยนั้นเป็นการพยากรณ์ค่าของตัวแปรหนึ่ง ซึ่งเรียกว่า ตัวแปรตามหรือตัวแปรเกณฑ์ (Dependent Variable) มักจะแทนด้วยตัว Y โดยใช้ค่าของตัวแปรอิสระหรือตัวแปรทำนาย (Independent Variable) ซึ่งแทนด้วยตัว X เป็นตัวทำนาย ตัวแปรทำนายนี้อาจใช้เพียงตัวเดียวหรือตั้งแต่ 2 ตัวขึ้นไป ถ้าใช้ตัวแปรทำนายเพียงตัวเดียวจะเรียกว่าการวิเคราะห์ถดถอยอย่างง่าย (Simple Regression Analysis) แต่ถ้าใช้ตัวแปรทำนายตั้งแต่ 2 ตัวขึ้นไป เรียกว่า การวิเคราะห์ถดถอยพหุคูณ (Multiple Regression Analysis) โดยทั่ว ๆ ไป การใช้ตัวแปรทำนายมากกว่าหนึ่งตัว ย่อมให้ผลในการพยากรณ์ถูกต้องมากกว่าการใช้ตัวแปรทำนายเพียงตัวเดียว (วิชิต หล่อจีระชุนห์กุล 2524 : 115)

ในการวิเคราะห์การถดถอยพหุคูณนี้ ผู้วิจัยต้องศึกษารูปแบบความสัมพันธ์ระหว่างตัวแปรทำนายกับตัวแปรเกณฑ์ก่อน โดยอาจจะมีความสัมพันธ์กันในลักษณะที่เป็นแบบเชิงเส้นตรง (linear) หรือไม่เป็นเชิงเส้นตรง (nonlinear) ก็ได้ ทั้งนี้ขึ้นอยู่กับข้อมูลที่นำมาศึกษาซึ่งการกำหนดลักษณะความสัมพันธ์ดังกล่าวมีความสำคัญมากต่อความถูกต้องในการพยากรณ์เนื่องจากค่าสัมประสิทธิ์สหสัมพันธ์พหุคูณที่คำนวนโดยวิธีการวิเคราะห์การถดถอยนั้น เป็นค่าสถิติที่แสดงถึงความสัมพันธ์ระหว่างตัวแปรเกณฑ์และผลรวมเชิงเส้นตรง (linear combination) ของตัวแปรทำนาย ค่าสถิติที่คำนวณได้จากกลุ่มตัวอย่าง จะเป็นตัวประมาณค่าที่ไม่เอนเอียงที่มี

ประสิทธิภาพสูงสุดในการประมาณค่าประชากร ก็ต่อเมื่อตัวแปรเกณฑ์และตัวแปรทำนายมีความสัมพันธ์กันเชิงเส้นตรงเท่านั้น (Blalock 1981 : 391) วิธีการวิเคราะห์ถดถอยที่นิยมใช้กันอย่างแพร่หลายก็คือ การวิเคราะห์ความถดถอยเชิงเส้น (Linear Regression Analysis) ซึ่งถ้ามีตัวแปรทำนายอยู่ p ตัว จะมีตัวแบบในการวิเคราะห์ดังนี้

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots\ldots + \beta_p X_p + E$$

เมื่อ Y = ตัวแปรเกณฑ์

β = สัมประสิทธิ์การถดถอย

X_i = ตัวแปรทำนาย เมื่อ $i = 1, 2, \ldots\ldots, P$

E = ค่าความแตกต่างระหว่างค่าประมาณที่ได้จากสมการกับค่าที่ได้จากการสังเกตจริง

จากสมการข้างต้น สามารถเขียนให้อยู่ในรูปของเมตริกซ์ (Matrix) ได้ดังนี้

$$\underset{(nx1)}{Y} = \underset{(nxp)}{X} \quad \underset{(px1)}{\beta} + \underset{(nx1)}{E}$$

เมื่อ n = ขนาดตัวอย่าง

p = จำนวนพารามิเตอร์สัมประสิทธิถดถอยเชิงเส้น

ตามปกติเมื่อผู้วิจัยนำเอาข้อมูลมาวิเคราะห์ความถดถอยเชิงเส้นจนกระทั่งภายหลังได้ตัวแบบถดถอยเชิงเส้น (Linear Regression Model) เพื่อนำไปใช้ในการพยากรณ์นั้นผู้วิจัยส่วนใหญ่ไม่ได้ทำการตรวจสอบความเหมาะสมของสมการถดถอยเชิงเส้น ว่ามีความเหมาะสมเพียงไหนก่อนที่จะนำเอาไปใช้ในการพยากรณ์จริง ๆ ในทางปฏิบัติจริงบางครั้งผู้วิจัยอาจมองข้ามตัวแปรที่สำคัญ ๆ บางตัวที่ต้องบรรจุเข้าไปในสมการถดถอย โดยเฉพาะตัวแปรทำนายในสมการถดถอย บางครั้งที่ผู้วิจัยอาจจะกำหนดตัวแปรทำนายน้อยเกินไป จนได้สมการถดถอยเชิงเส้นไม่เหมาะสมที่จะใช้ในการพยากรณ์และบางครั้งผู้วิจัยอาจจะกำหนดตัวแปรทำนายมากเกิน

ไปจนไม่ทราบว่าจะ เลือกใช้ตัวแปรทำนายตัวไหนดีจึงจะทำให้ได้สมการถดถอย เชิง เส้นที่ดี เพื่อใช้ในการพยากรณ์

ในการกำหนดตัวแปรทำนาย เพื่อใช้ในสมการถดถอย เชิง เส้นในการพยากรณ์นั้นถ้าใช้ตัวแปรทำนาย เพียงตัว เดียว อาจจะทำให้การพยากรณ์คลาดเคลื่อนจากความเป็นจริงหรืออาจจะ เชื่อถือไม่ได้มากนัก เป็นเพราะว่าตัวแปรทำนายนั้นไม่ดี เท่าที่ควร หรือตัวแปรทำนายที่อยู่ในสมการนั้นไม่ดี เท่าตัวแปรทำนายที่ไม่ได้นำมาวิ เคราะห์ ดังนั้นในการวิ เคราะห์ความถดถอยอย่างง่าย (Simple Linear Regression) จึงไม่เป็นที่นิยมใช้สำหรับการพยากรณ์ ส่วนใหญ่แล้วนักวิจัยจะใช้ตัวแปรทำนายหลายตัวในการวิ เคราะห์ คือ การวิ เคราะห์ถดถอยพหุคูณ เชิง เส้น (Multiple Linear Regression) ซึ่งจะช่วยให้การพยากรณ์แม่นยำและถูกต้องมากขึ้น แต่ในการใช้ตัวแปรทำนายมาก เกินไปนั้นไม่ใช่ว่าจะให้ผลดี เสมอไป บางครั้งก็ไม่ช่วยลดความคลาด เคลื่อนในการพยากรณ์ได้เลย การ เพิ่มหรือลดความคลาด เคลื่อนในการพยากรณ์นั้นขึ้นอยู่กับตัวแปรทำนายที่ เข้าไปอยู่ในสมการถดถอยนั้น เหมาะสม เพียงไหน ดังนั้นจึงจำ เป็นจะต้องมีการคัด เลือกตัวแปรทำนายที่มีอยู่มากมายขึ้น เข้าไปในสมการถดถอย เชิง เส้นเพื่อให้ได้สมการที่ดีหรือ เหมาะสมที่สุดเพื่อใช้ในการพยากรณ์ วิธีคัด เลือกตัวแปรทำนาย เข้าสู่สมการถดถอยมีอยู่หลายวิธีด้วยกัน แต่วิธีที่นิยมกันแพร่หลายในปัจจุบันก็คือ การ เลือกตัวแปรแบบไปข้างหน้า (Forward Selection) การกำจัดตัวแปรแบบถอยหลัง (Backward Elimination) (Efroymson 1966 or Draper and smith 1966) และการถดถอยแบบขั้นบันได (Stepwise Regression) (Efroymson 1960) ซึ่งในแต่ละวิธีจะมีหลัก เกณฑ์และวิธีการที่แตกต่างกัน และในการวิ เคราะห์ถดถอยพหุคูณนั้น ยังมีปัญหาที่สำคัญ นั่นคือ ระดับความสัมพันธ์ระหว่างตัวแปรทำนายด้วยกัน เอง ซึ่งถ้าตัวแปรทำนายมีความสัมพันธ์กันอยู่ในระดับต่ำก็จะสามารถร่วมกันอธิบายตัวแปร เกณฑ์ได้ดี แต่ถ้าตัวแปรทำนายมีความสัมพันธ์กันมากขึ้นความสามารถที่จะร่วมกันอธิบายตัวแปร เกณฑ์จะลดลง นอกจากนี้ยังมีข้อจำกัด เกี่ยวกับขนาดของกลุ่มตัวอย่างต่อจำนวนตัวแปรทำนายซึ่งจะมีผลต่อการลดลงของค่าสหสัมพันธ์พหุคูณกำลังสอง (R^2) เมื่อนำไปใช้กับกลุ่มตัวอย่างใหม่ ซึ่ง เป็นปัญหาที่พบอยู่ เสมอในงานวิจัยทางด้านสังคมศาสตร์ ซึ่งมีรายละ เอียดดังนี้

1. ผลที่เกิดจากตัวแปรทำนายมีสหสัมพันธ์ระหว่างกันสูง (Multicollinearity) จะทำให้เกิดปัญหาในด้านการแปลความหมายของข้อมูลทำให้ความคลาดเคลื่อนมาตราฐานของสัมประสิทธ์ถดถอยพหุ (standard error of b) มีค่ามาก นั่นหมายถึงว่าค่าประมาณสัมประสิทธิ์ถดถอยพหุที่ได้ขาดความแม่นตรง

จากสมการ

$$S_{by1.2...k} = \sqrt{\frac{S^2_{y.12...k}}{\Sigma X^2_1 \ (1-R^2_{1.2...k})}}$$ (Pedhazur 1973 : 119)

เมื่อ

$S_{by1.2...k}$ = ความคลาดเคลื่อนมาตราฐานของ b_1

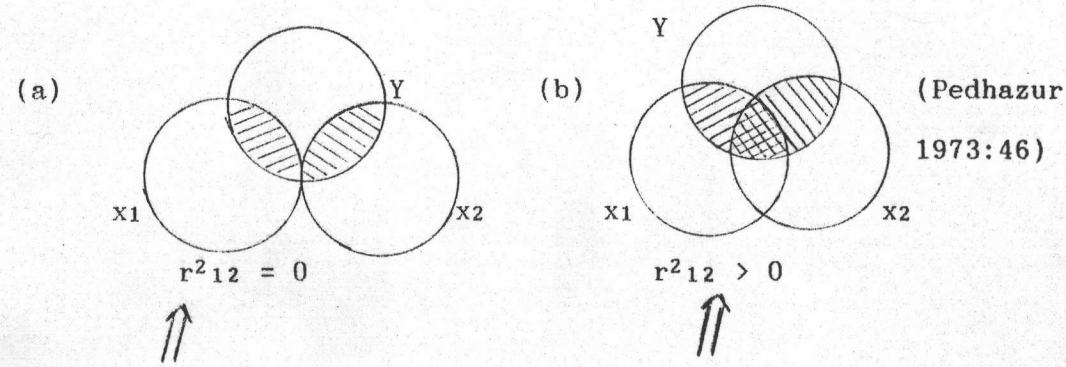
$S^2_{y.12...k}$ = ความแปรปรวนของการประมาณด้วยตัวแปรทำนาย

ΣX^2_1 = $\Sigma (X_1 - \bar{X}_1)^2$

$R_{1.2...k}$ = ค่าสหสัมพันธ์พหุคูณระหว่าง X_1 ซึ่งถือเสมือนตัวแปรเกณฑ์ โดย X_2 ถึง X_k เป็นตัวแปรทำนาย

จะเห็นว่าความคลาดเคลื่อนมาตราฐานของการประมาณค่า b_1 จะมีมากหรือน้อยขึ้นอยู่กับขนาดของสหสัมพันธ์พหุคูณระหว่างตัวแปรทำนาย ($R_{1.2...k}$) ถ้าสหสัมพันธ์พหุคูณระหว่างตัวแปรทำนายสูงก็จะทำให้ความคลาดเคลื่อนมาตราฐานของสัมประสิทธิ์ถดถอยมีขนาดใหญ่ นอกจากนี้ยังทำให้ค่าสัมประสิทธิ์สหสัมพันธ์ส่วนย่อย (Partial Correlation Coeffieient) ที่ใช้สำหรับการคัดเลือกตัวแปรทำนาย เข้าสู่สมการถดถอยเกิดความคลาดเคลื่อนอีกด้วย ซึ่งจะส่งผลต่อลำดับที่ของการคัดตัวแปรเข้าสมการ (สุชาติ ประสิทธิรัฐสินธุ์ 2522 : 186) จะทำให้ลำบากในการเปรียบเทียบความสำคัญของตัวแปรทำนายที่มีต่อตัวแปรเกณฑ์เพราะลำดับที่ของการเข้าสู่สมการของตัวแปรทำนายนั้น ย่อมมีผลกระทบต่อลำดับความสำคัญที่มีต่อการอธิบายความแปรปรวนของตัวแปรเกณฑ์ เพราะตัวแปรทำนายที่เข้าสู่สมการในอันดับหลังย่อมไปทับซ้อนกับผลกระทบของตัวแปรตัวแรกที่มีต่อตัวแปรเกณฑ์ (ปุระชัย เปี่ยมสมบูรณ์ 2527 : 30) และยังทำให้อธิบายค่าความแปรปรวนสะสมที่เพิ่มขึ้นจากเดิมได้ค่อนข้างต่ำอีกด้วย

แผนภาพที่ 1 แสดงความสัมพันธ์ระหว่างตัวแปรทำนายที่ร่วมกันอธิบายความแปรปรวนที่มีต่อตัวแปรเกณฑ์



$R^2_{y.12} = r^2_{y.1} + r^2_{y.2}$ $R^2_{y.12} = r^2(X_1 \cap Y) + r^2(X_2 \cap Y) - r^2(X_1 \cap X_2 \cap Y)$

จะเห็นว่าจาก (a) ตัวแปรทำนาย X_1 กับ X_2 ไม่มีความสัมพันธ์กัน ($r_{12} = 0$) ฉะนั้นจึงร่วมกันอธิบายความแปรปรวนที่มีต่อตัวแปรเกณฑ์ (Y) ได้สูงดังสมการที่แสดงไว้ จาก(b) ตัวแปรทำนาย X_1 กับ X_2 มีความสัมพันธ์กับ Y เท่า ๆ กัน ฉะนั้นโดยลำพังแต่ละตัวจะมีกำลังการอธิบายความแปรปรวนที่มีต่อ Y ได้ไม่แตกต่างกัน แต่ X_1 กับ X_2 มีความสัมพันธ์ต่อกันสูง ถ้า X_1 ถูกคัดเข้าสมการก่อน และ X_2 ถูกคัดเข้าสมการทีหลัง X_2 จะสามารถช่วยอธิบายความแปรปรวนของ Y ได้เพิ่มขึ้นน้อยมากเมื่อเทียบกับค่าความแปรปรวนของ Y ที่ถูกอธิบายโดย X_1 เพราะต้องนำส่วนที่ทับซ้อนนั้นมาหักออกดังสมการที่แสดงไว้ และถ้ามีตัวแปรทำนายหลายตัว และตัวแปรเหล่านั้นมีความสัมพันธ์ต่อกันในระดับค่อนข้างสูง ตัวแปรทำนายที่ถูกคัดเข้าสมการในลำดับที่ 3,4 และต่อ ๆ ไป จะยิ่งเพิ่มค่าความแปรปรวนสะสมของ Y ได้น้อยลงตามลำดับฉะนั้นจะเห็นว่า ถ้าตัวแปรทำนายมีความสัมพันธ์ระหว่างกันสูง (Multicollinearity) จะส่งผลต่อค่าความแปรปรวนที่มีต่อตัวแปรเกณฑ์ดังกล่าว

2. ผลของขนาดกลุ่มตัวอย่าง (Sample Size) ต่อจำนวนตัวแปรทำนายที่มีต่อความคลาดเคลื่อนมาตราฐานของการประมาณ (The Standard error of Estimate) ในการสร้างสมการถดถอยนั้นจะต้องใช้ข้อมูลจำนวนมาก ทั้งนี้จากการคำนวนความคลาดเคลื่อนมาตราฐานของการประมาณ

$$S_{y.12...p} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}$$

เมื่อ

$S_{y.12...p}$ = ความคลาดเคลื่อนมาตราฐานของการประมาณค่า

y_i = ตัวแปรเกณฑ์ที่ได้จากการสังเกตจริง

$\hat{y}_i$ = ผลรวมเชิงเส้นตรงของตัวแปรทำนาย (Linear Combination)

N = ขนาดของกลุ่มตัวอย่าง

p = จำนวนตัวแปรทำนาย

จะเห็นว่าความคลาดเคลื่อนมาตราฐานของการประมาณจะมีมากหรือน้อยขึ้นอยู่กับขนาดของกลุ่มตัวอย่าง ซึ่งถ้าหากผู้วิจัยใช้กลุ่มตัวอย่างที่มีขนาดเล็กแล้วจะทำให้ความคลาดเคลื่อนมีขนาดใหญ่ ทำให้เกิดปัญหาในการทดสอบสมมุติฐานเพื่ออ้างอิงไปสู่ประชากรและการตัดสินใจว่าตัวแบบสมการที่สร้างขึ้นจากกลุ่มตัวอย่างที่ศึกษาซึ่งประกอบด้วยตัวแปรพยากรณ์ (Xs) นั้นมีความสัมพันธ์กับตัวแปรเกณฑ์ (Y) อย่างมีนัยสำคัญหรือไม่ ดังจะเห็นได้จากปัญหาการลดลงของค่าสัมประสิทธิ์สหสัมพันธ์พหุคูณยกกำลังสอง (Shrinkage) เมื่อผู้วิจัยนำสมการถดถอยที่สร้างขึ้นจากกลุ่มตัวอย่างหนึ่งไปใช้พยากรณ์ในอีกกลุ่มตัวอย่างหนึ่งที่สุ่มมาจากประชากรเดียวกัน (Pedhazur 1982:147-149)

จากสมการ

$$\hat{R}^2 = 1-(1-R^2)\frac{(N-1)}{(N-k-1)}$$

เมื่อ

$\hat{R}^2$ = สหสัมพันธ์พหุคูณกำลังสองของการประมาณของประชากร

R^2 = สหสัมพันธ์พหุคูณกำลังสองที่คำนวนได้จากกลุ่มตัวอย่าง

N = ขนาดของกลุ่มตัวอย่าง

k = จำนวนตัวแปรอิสระ

ถ้ามีตัวแปรทำนายอยู่ 3 ตัว ค่าสหสัมพันธ์พหุคูณกำลังสองระหว่างตัวแปรทำนายและตัวแปรเกณฑ์ (R^2) มีค่าเท่ากับ .36 ค่า $\hat{R}^2$ จะเป็นเท่าใด ถ้าขนาดของกลุ่มตัวอย่างต่อจำนวนตัวแปรทำนายเป็น 5:1, 30:1 และ 50:1

ขนาดของกลุ่มตัวอย่าง : จำนวนตัวแปรทำนาย = 5:1

$$\hat{R}^2 = 1-(1-.36)\frac{(15-1)}{(15-3-1)} = .19$$

ขนาดของกลุ่มตัวอย่าง : จำนวนตัวแปรทำนาย = 30:1

$$\hat{R}^2 = 1-(1-.36)\frac{(90-1)}{(90-3-1)} = .34$$

ขนาดของกลุ่มตัวอย่าง : จำนวนตัวแปรทำนาย = 50:1

$$\hat{R}^2 = 1-(1-.36)\frac{(150-1)}{(150-3-1)} = .35$$

จะเห็นว่า ยิ่งขนาดของกลุ่มตัวอย่างมากเท่าใด จะทำให้ค่า R^2 และ $\hat{R}^2$ ใกล้เคียงกันมากขึ้นเท่านั้น นอกจากนั้น Halinski และ Feldt (1970) ได้เสนออัตราส่วนต่ำสุดระหว่างขนาดของกลุ่มตัวอย่างต่อจำนวนตัวแปรทำนายที่ใช้ในสมการถดถอยโดยการซิมูเลชั่น (Simulation) นั้นควรใช้ 10:1 ในขณะเดียวกัน Miller และ Kunce (1973) ได้เสนอแนะไว้ด้วยอัตราส่วนที่เท่ากัน ส่วน Lindeman (1980) ได้เสนอว่าในการวิเคราะห์สหสัมพันธ์พหุคูณนั้นควรใช้ 20:1 ดังจะเห็นว่าผู้วิจัยได้ให้ข้อเสนอแนะที่แตกต่างกันออกไปในการวิจัยครั้งนี้ ผู้วิจัยจึงเลือกใช้ขนาดของกลุ่มตัวอย่างต่อจำนวนตัวแปรทำนายเป็น 10:1 ตามที่ Halinski และ Feldt เสนอไว้ เพราะข้อมูลในการวิจัยครั้งนี้สร้างขึ้นด้วยการซีมูเลชั่น (Simulation) ส่วนจำนวนตัวแปรทั้งหมดที่ศึกษา เพื่อให้สามารถใช้สูตรในการสร้างข้อมูลได้อย่างมีประสิทธิภาพ คือ มีความคลาดเคลื่อนของข้อมูลต่ำที่สุด ผู้วิจัยจึงเลือกศึกษากับจำนวนตัวแปรทั้งหมด 6 ตัว

เนื่องจากการวิเคราะห์ถดถอยพหุคูณ มีเทคนิควิธีที่ใช้ในการคัดเลือกตัวแปรทำนายอยู่หลายวิธี แต่วิธีที่นิยมใช้กันอย่างแพร่หลายในปัจจุบันคือการเลือกตัวแปรแบบไปข้างหน้า (Forward Selection) การกำจัดตัวแปรแบบถอยหลัง (Backward Elimination) และการถดถอยแบบขั้นบันได (Stepwise Regression) ซึ่งการเลือกตัวแปรแบบไปข้างหน้ามีข้อเสียตรงที่ไม่ได้มีการพิจารณาบทบาทของตัวแปรทำนายที่รวมอยู่ในสมการในขั้นก่อน เมื่อมี

ตัวแปรตัวใหม่เข้าไปอยู่สมการ การกำจัดตัวแปรแบบถอยหลังดีตรงที่ได้พิจารณาตัวแปรทำนายทุกตัวในตอนเริ่มแรก แต่มีข้อเสียเพราะใช้เวลาในการคำนวนมากและไม่ได้นำตัวแปรทำนายที่ออกจากสมการไปแล้วนั้นกลับเข้ามาพิจารณาอีก ส่วนการถดถอยแบบขั้นบันไดมีลักษณะผสมผสานระหว่างวิธีการเลือกตัวแปรแบบไปข้างหน้า และการกำจัดตัวแปรแบบถอยหลัง โดยที่ในแต่ละครั้งที่เพิ่มตัวแปรทำนายใหม่เข้าไปในสมการ จะนำเอาตัวแปรทำนายที่เข้าสู่สมการแล้วมาพิจารณาใหม่ว่ายังมีความสำคัญต่อการอธิบายค่าความแปรปรวนของตัวแปรเกณฑ์อีกต่อไปหรือไม่ แต่อย่างไรก็ตามทั้ง 3 วิธีนี้มีข้อเสียมากในกรณีที่ตัวแปรทำนายมีความสัมพันธ์ระหว่างกันสูง (Multicollinearty) ซึ่งจะส่งผลต่อค่าความแปรปรวนที่มีต่อตัวแปรเกณฑ์อย่างมาก ซึ่งค่าความแปรปรวนนี้อธิบายโดยค่าสหสัมพันธ์พหุคูณกำลังสอง หรือค่าสัมประสิทธิ์การทำนาย (Coefficient of Determination หรือ R^2) จึงทำให้เกิดปัญหาแก่ผู้วิจัยเสมอมา ในการเลือกใช้เทคนิคการคัดเลือกตัวแปรทำนายว่าควรจะเลือกใช้วิธีใดในสถานการณ์ที่มีความสัมพันธ์ระหว่างตัวแปรทำนายระดับใด จึงจะให้ผลดีที่สุด ในการคัดเลือกตัวแปรทำนายเข้าสู่สมการถดถอยเชิงเส้น จากเหตุผลดังกล่าวผู้วิจัยจึงจะศึกษาเปรียบเทียบผลการวิเคราะห์ที่ได้จากการใช้เทคนิคการคัดเลือกตัวแปรทำนายเข้าสู่สมการถดถอยใน 3 วิธี คือ การเลือกตัวแปรแบบไปข้างหน้า (Forward Selection) การกำจัดตัวแปรแบบถอยหลัง (Backward Elimination) และการถดถอยแบบขั้นบันได (Stepwise Regression) เมื่อระดับความสัมพันธ์ระหว่างตัวแปรทำนายแตกต่างกัน โดยเปรียบเทียบจากค่าเฉลี่ยและความแปรปรวนของสัมประสิทธิ์สหสัมพันธ์พหุคูณกำลังสอง (R^2) เพื่อที่จะได้ทราบว่าความสัมพันธ์ระหว่างตัวแปรทำนายที่ทำให้ได้ค่าเฉลี่ยของสัมประสิทธิ์สหสัมพันธ์พหุคูณกำลังสอง (R^2) สูงสุด และมีความแปรปรวนต่ำสุด เพราะนั่นหมายความว่า กลุ่มของตัวแปรทำนายที่ถูกคัดเลือกเข้าสมการภายใต้สถานการณ์นั้น ๆ สามารถพยากรณ์ตัวแปรเกณฑ์ได้ดีที่สุด และเพื่อเปรียบเทียบร้อยละของตัวแปรทำนายที่ถูกคัดเข้าสมการในอันดับต่าง ๆ ด้วยวิธีการเลือกตัวแปรแบบไปข้างหน้า (Forward Selection) และวิธีถดถอยแบบขั้นบันได (Stepwise Regression) เมื่อตัวแปรทำนายมีความสัมพันธ์ในระดับเดียวกัน

วัตถุประสงค์ของการวิจัย

เพื่อศึกษาเปรียบเทียบค่าสถิติที่ได้จากเทคนิควิธีการคัดเลือกตัวแปรทำนายเข้าสู่สมการถดถอย 3 วิธี คือ การเลือกตัวแปรแบบไปข้างหน้า (Forward Selection) การกำจัดตัวแปรแบบถอยหลัง (Backward Elimination) และการถดถอยแบบขั้นบันได (Stepwise Regression) ดังนี้

1. เปรียบเทียบค่าเฉลี่ยและความแปรปรวนของสัมประสิทธิ์สหสัมพันธ์พหุคูณกำลังสอง ที่ได้จากการคัดเลือกตัวแปรทั้ง 3 วิธี เมื่อความสัมพันธ์ระหว่างตัวแปรทำนายอยู่ในระดับเดียวกัน

2. เปรียบเทียบค่าเฉลี่ย และความแปรปรวนของสัมประสิทธิ์สหสัมพันธ์พหุคูณกำลังสอง (R^2) ที่ได้จากการคัดเลือกตัวแปรวิธีเดียวกัน เมื่อความสัมพันธ์ระหว่างตัวแปรทำนายต่างกัน

3. เปรียบเทียบร้อยละของตัวแปรทำนายที่ถูกคัดเข้าสมการในอันดับต่าง ๆ ด้วยวิธีการเลือกตัวแปรแบบไปข้างหน้า (Forward Selection) และ วิธีถดถอยแบบขั้นบันได (Stepwise Regression) เมื่อความสัมพันธ์ระหว่างตัวแปรทำนายอยู่ในระดับเดียวกัน

ภายใต้เงื่อนไขต่อไปนี้

1. ความสัมพันธ์ระหว่างตัวแปรทำนายแบ่งออกเป็น 3 ระดับ มีค่าอยู่ระหว่าง (0.00-0.30), (0.30-0.70) และ (0.70-1.00)
2. ความสัมพันธ์ระหว่างตัวแปรทำนายกับตัวแปรเกณฑ์มีค่าอยู่ระหว่าง (0.30-1.00)
3. จำนวนตัวแปรทำนาย 5 ตัว
4. ขนาดของกลุ่มตัวอย่างเป็น 10 เท่าของตัวแปรทำนาย

สมมุติฐานการวิจัย

ผลของการวิเคราะห์ถดถอยพหุคูณด้วยการใช้เทคนิควิธีทั้ง 3 วิธีในการคัดเลือกตัวแปรทำนาย เมื่อความสัมพันธ์ระหว่างตัวแปรทำนายอยู่ในระดับเดียวกัน น่าจะได้ค่าเฉลี่ยและความแปรปรวนของสัมประสิทธิ์สหสัมพันธ์พหุคูณกำลังสอง (R^2) แตกต่างกัน

เนื่องจากวิธีการคัดเลือกตัวแปรแต่ละวิธีมีขั้นตอนในการคัดตัวแปรต่างกัน วิธีการเลือกตัวแปรแบบไปข้างหน้า (Forward Selection) จะคัดตัวแปรเข้าทีละตัว โดยพิจารณาจากค่า R^2 ที่เพิ่มขึ้น แต่มีข้อเสียตรงที่ไม่ได้นำเอาตัวแปรที่เข้าสมการไปแล้วมาพิจารณาอีก เมื่อมีตัวแปรตัวใหม่เข้าไปอยู่ในสมการ วิธีการกำจัดตัวแปรแบบถอยหลัง (Backward Elimination) จะคัดตัวแปรออกมาทีละตัว โดยพิจารณาจากค่า R^2 ที่ลดลง แต่ไม่ได้นำเอาตัวแปรที่ออกจากสมการแล้วมาพิจารณาอีก เมื่อจะคัดตัวแปรตัวใหม่ออกจากสมการ ส่วนวิธีถดถอยแบบขั้นบันได (Stepwise Regression) จะปิดข้อบกพร่องของทั้งสองวิธีข้างต้น เพราะทุกครั้งที่จะเพิ่มตัวแปรตัวใหม่เข้าไปในสมการ จำนำเอาตัวแปรเดิมที่อยู่ในสมการแล้วมาพิจารณาใหม่ และตัวแปรบางตัวที่ถูกคัดออกจากสมการก็นำมาพิจารณาใหม่อีกทุกครั้งที่จะคัดตัวแปรใหม่เข้าสมการ เมื่อแต่ละวิธีมีวิธีการคัดตัวแปรทำนายต่างกันเช่นนี้ ตัวแปรที่ถูกคัดเข้าสมการจึงแตกต่างกันในแต่ละวิธี ฉะนั้นค่าเฉลี่ยและความแปรปรวนของสัมประสิทธิ์สหสัมพันธ์พหุคูณกำลังสอง (R^2) ที่ได้จึงน่าจะแตกต่างกัน

ขอบเขตการวิจัย

1. ตัวแปรที่ศึกษา

ก. ตัวแปรต้น (Independent Variable) คือ

1.1 เทคนิควิธีการคัดเลือกตัวแปรทำนายเข้าสู่สมการถดถอยทั้ง 3 วิธี ได้แก่

1.1.1 การเลือกตัวแปรแบบไปข้างหน้า (Forward Selection)

1.1.2 การกำจัดตัวแปรแบบถอยหลัง (Backward Elimination)

1.1.3 การถดถอยแบบขั้นบันได (Stepwise Regression)

1.2 ระดับความสัมพันธ์ระหว่างตัวแปรทำนาย แบ่งออกเป็น 3 ระดับคือ มีค่าอยู่ระหว่าง (0.00-0.30), (0.30-0.70), (0.70-1.00)

หอสมุดกลาง สถาบันวิทยบริการ จุฬาลงกรณ์มหาวิทยาลัย

1.3 ระดับความสัมพันธ์ระหว่างตัวแปรทำนายกับตัวแปรเกณฑ์ มีค่าอยู่ระหว่าง (0.30-1.00)

1.4 จำนวนตัวแปรทำนาย 5 ตัว

1.5 ขนาดของกลุ่มตัวอย่าง : จำนวนตัวแปรทำนาย = 10 : 1

ข. ตัวแปรตาม (Dependent Variable) คือค่าเฉลี่ยและความแปรปรวนของค่าสัมประสิทธิ์สหสัมพันธ์พหุคูณกำลังสอง (R^2) ที่คำนวนได้จากวิธีการเลือกตัวแปรแบบไปข้างหน้า การกำจัดตัวแปรแบบถอยหลัง และการถดถอยแบบขั้นบันได

2. ข้อมูลประชากรที่นำมาศึกษามีการแจกแจงแบบปกติหลายตัวแปร (Multivariate normal distribution) $X \sim N(\mu, \Sigma)$ โดยที่

$$\mu = 0$$

$$\Sigma = \sigma_{ij}, \sigma_{ij} = \sigma \text{ เมื่อ } i \neq j$$

$$= 1 \text{ เมื่อ } i = j$$

3. แบ่งระดับความสัมพันธ์ระหว่างตัวแปรทำนายออกเป็น 3 ระดับ มีค่าอยู่ระหว่าง (0.00-0.30), (0.30-0.70), (0.70-1.00)

4. ความสัมพันธ์ระหว่างตัวแปรทำนายกับตัวแปรเกณฑ์ มีค่าอยู่ระหว่าง 0.30-1.00

<u>ความไม่สมบูรณ์ของการวิจัย</u>

1. การวิจัยครั้งนี้ได้กำหนดให้มีจำนวนตัวแปรทำนายและขนาดของกลุ่มตัวอย่างคงที่ จึงทำให้ผลงานวิจัยนำไปใช้ได้ในวงจำกัด ถ้าจะให้สมบูรณ์ควรจะกำหนดให้มีจำนวนตัวแปรทำนายและขนาดของกลุ่มตัวอย่างหลาย ๆ ขนาด

2. ในการสร้างความสัมพันธ์ระหว่างตัวแปรทำนายและระหว่างตัวแปรทำนายกับตัวแปรเกณฑ์ ได้ใช้วิธีสุ่ม (Random) เพื่อให้ได้ความสัมพันธ์ตามระดับที่กำหนด จึงทำให้ได้เฉพาะค่าบางค่าเท่านั้น ซึ่งจะไม่ครอบคลุมค่าของความสัมพันธ์ทั้งหมด ถึงแม้ว่าจะอยู่ในช่วงที่กำหนดก็ตาม

ความจำกัดของการวิจัย

1. ข้อมูลที่ใช้ในการวิจัยครั้งนี้ได้จากการจำลองด้วยเทคนิคมอนติคาร์โลซิมูเลชั่น (Monte Carlo Simulation) ฉะนั้นผลงานวิจัยจึงนำไปใช้ได้กับงานวิจัยที่เก็บข้อมูลจากกลุ่มตัวอย่างจริง เฉพาะกรณีที่ข้อมูลนั้นได้เกิดขึ้นตรงตามเงื่อนไขที่กำหนดในงานวิจัยครั้งนี้เท่านั้น

2. การใช้โปรแกรม $SPSS_x$ ในการประมวลผล เพื่อศึกษาอันดับที่ของตัวแปรทำนายที่ถูกคัดเลือกเข้าสมการนั้น ผลที่พิมพ์ออกมา (print out) จะแสดงอันดับที่ของตัวแปรที่ถูกคัดเข้าสมการทุกชุดเฉพาะวิธีคัดเลือกตัวแปรแบบไปข้างหน้า (Forward Selection) และวิธีถดถอยแบบขั้นบันได (Stepwise Regression) เท่านั้น ส่วนวิธีกำจัดตัวแปรแบบถอยหลัง (Backward Elimination) จะแสดงผลเฉพาะชุดที่มีตัวแปรบางตัวถูกคัดออกเท่านั้น ฉะนั้นในการเปรียบเทียบร้อยละของตัวแปรทำนายที่ถูกคัดเข้าสมการในอันดับต่าง ๆ นั้น จะเปรียบเทียบได้เฉพาะ 2 วิธีข้างต้น ดังกล่าวเท่านั้น

ข้อตกลงเบื้องต้น

ข้อมูลที่ใช้ในการวิจัยครั้งนี้สร้างขึ้นจากโปรแกรมฟังก์ชั่นที่มีหลักการศึกษาว่าเลขสุ่มที่ผลิตได้มีลักษณะสุ่ม (Random) จริง แต่ละหมายเลขมีโอกาสเกิดขึ้นเท่าเทียมกันและมีจำนวนเพียงพอที่จะใช้ในการศึกษา

คำจำกัดความในการวิจัย

ค่าสัมประสิทธิ์สหสัมพันธ์พหุคูณ (R) หมายถึง ค่าที่บอกถึงระดับความสัมพันธ์ระหว่างฟังก์ชั่นผลรวมเชิงเส้นตรงของตัวแปรทำนายกับตัวแปรเกณฑ์

ค่าสัมประสิทธิ์สหสัมพันธ์พหุคูณกำลังสอง (R^2) หมายถึง ค่าที่บอกถึงสัดส่วนของความแปรปรวนของตัวแปรเกณฑ์ที่ลดลง ถ้า R^2 มีค่ามาก ก็จะบอกได้ว่าความแปรปรวนของตัวแปรเกณฑ์ลดลงมาก นั่นคือค่าของตัวแปรเกณฑ์ที่ได้จากการสังเกตกับค่าที่ได้จากการทำนายมีค่าใกล้เคียงกัน

ซิมูเลชั่น (Simlution) หมายถึง เทคนิคการผลิตเลขสุ่ม (Technique of random number generation) เลขสุ่มที่ได้จะมีลักษณะสุ่ม (random) แต่ละหมายเลขจะมีโอกาสเกิดขึ้นได้เท่าเทียมกัน

ความแปรปรวน หมายถึง ความแตกต่างของค่าที่ได้จากการสังเกตกับค่าเฉลี่ยของตัวแปรเกณฑ์

เทคนิควิธีการคัดเลือกตัวแปร 3 วิธี หมายถึง วิธีการคัดเลือกตัวแปรทำนายเข้าสู่สมการถดถอย ได้แก่ วิธีเลือกตัวแปรแบบไปข้างหน้า (Forward Selection) วิธีกำจัดตัวแปรแบบถอยหลัง (Backward Elimination) และวิธีถดถอยแบบขั้นบันได (Stepwise Regression)

ประโยชน์ที่คาดว่าจะได้รับ

1. เพื่อให้ทราบวิธีที่ดีที่สุดที่ใช้สำหรับการคัดเลือกตัวแปรทำนายเข้าสู่สมการถดถอย เมื่อมีระดับความสัมพันธ์ระหว่างตัวแปรทำนาย ระดับความสัมพันธ์ระหว่างตัวแปรทำนายกับตัวแปรเกณฑ์ จำนวนตัวแปรทำนาย และขนาดของกลุ่มตัวอย่างต่อจำนวนตัวแปรทำนาย ตามที่กำหนดในงานวิจัยนี้

2. เพื่อให้นักวิจัยนำผลสรุปที่ได้ไปใช้ในการเลือกวิธีการคัดเลือกตัวแปรทำนายเข้าสู่สมการถดถอยได้อย่างเหมาะสมกับข้อมูล

3. เพื่อเป็นพื้นฐานในการศึกษาวิจัยด้านนี้ต่อไป