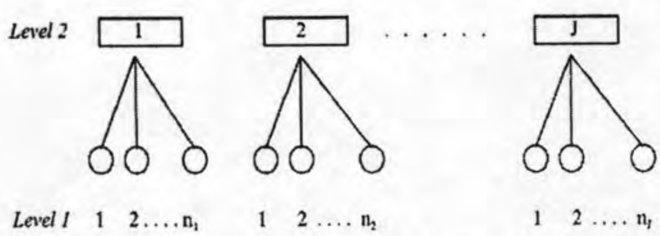
บทที่ 1

บทนำ

ความเป็นมาและความสำคัญของปัญหา

การวิเคราะห์ความถดถอยเชิงเส้น (Linear Regression Analysis) เป็นวิธีที่ใช้กันอย่างแพร่หลายในงานวิจัยด้านต่างๆทั้งในด้านวิศวกรรม เศรษฐศาสตร์ วิทยาศาสตร์ อุตสาหกรรมและในทางสังคมศาสตร์ เป็นต้น การวิเคราะห์นี้เป็นวิธีการทางสถิติที่นำมาใช้เพื่อศึกษาหาความสัมพันธ์ระหว่างตัวแปรตาม (Dependent Variable) และตัวแปรอิสระ (Independent Variables) ซึ่งอาจมีเพียงหนึ่งตัวหรือมากกว่าก็ได้ โดยทั่วไปแล้วในการวิจัยทางสังคมศาสตร์ ทางการศึกษา หรือทางการแพทย์ ผู้วิจัยมักพบข้อมูลที่ลักษณะติดกลุ่ม (Nested) หรือมีโครงสร้างในลักษณะเป็นเชิงลำดับชั้น (Hierarchical Structure) เช่น ข้อมูลของนักเรียนในแต่ละโรงเรียน ข้อมูลพนักงานในแต่ละองค์กร ข้อมูลของครัวเรือนในแต่ละจังหวัด หรือข้อมูลคนไข้ที่ได้รับการรักษาจากแพทย์แต่ละคน เป็นต้น จากตัวอย่างดังกล่าวเป็นข้อมูลที่มีโครงสร้างเป็นเชิงลำดับชั้นที่มี 2 ระดับ (Two-Level Hierarchies) โดยสามารถเขียนเป็นแผนภาพได้ดังแสดงในรูปที่ 1.1 โครงสร้างของข้อมูลดังกล่าวอาจมีมากกว่า 2 ระดับ



ภาพประกอบที่ 1.1 แสดงลักษณะโครงสร้างข้อมูลเชิงลำดับชั้นที่มี 2 ระดับ

ได้ เช่น นักเรียนติดกลุ่มในห้องเรียน และห้องเรียนติดกลุ่มในโรงเรียนซึ่งมีอยู่หลายโรงเรียน ลักษณะเช่นนี้เรียกว่าข้อมูลที่มีโครงสร้างเป็นเชิงลำดับชั้นที่มี 3 ระดับ (Three-Level Hierarchies) ซึ่งในการวิเคราะห์ก็จะมีความยุ่งยากซับซ้อนมากขึ้นตามลักษณะข้อมูลที่ซับซ้อนขึ้น ในการวิจัยต่อไปนี้จะกล่าวถึงแต่ในกรณีที่ข้อมูลมีโครงสร้างเป็นเชิงลำดับที่มี 2 ระดับเท่านั้น โดยหากข้อมูลมีโครงสร้างในลักษณะเชิงลำดับชั้นดังที่กล่าวมา แล้วผู้วิจัยละเลยโครงสร้างของข้อมูลโดยยังใช้การวิเคราะห์การถดถอยแบบปกติซึ่งเป็นการวิเคราะห์แบบระดับเดียว (Single-Level Analysis) ในการวิเคราะห์จะทำให้เกิดความผิดพลาดขึ้น โดยจะทำให้การประมาณค่าคลาดเคลื่อนมาตรฐานผิดพลาด (Misestimated Standard Error) ส่งผลให้การอนุมานเกี่ยวกับค่าพารามิเตอร์ในตัวแบบ

มีความผิดพลาดเพิ่มมากขึ้นนอกจากนี้การละเลยโครงสร้างของข้อมูลยังทำให้ ผู้วิจัยไม่สามารถวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรที่อยู่ต่างระดับกันได้ ทำให้เกิดความเอนเอียงในการสรุปผลระหว่างระดับ (Aggregation bias) อีกทั้งมีโอกาสที่ความสัมพันธ์ของตัวแปรตาม และตัวแปรอิสระที่อยู่ต่างกลุ่มกันจะมีความแตกต่างกัน (Heterogeneity of Regression Coefficients) ซึ่งทำให้ค่าสัมประสิทธิ์ถดถอยที่ได้จากการวิเคราะห์ข้อมูลระดับเดียวนั้นมีความผิดพลาด จากปัญหาต่างๆดังที่กล่าวมานั้นจะนำไปสู่ข้อสรุปที่ผิดพลาดในงานวิจัยได้ (ศิริชัย กาญจนวาสี, 2548 : 85-86; Goldstein, 1995 :1-10; Snijder and Bosker, 1999 :13-16, Raudenbush and Bryk; 2002 : 5)

การวิเคราะห์พหุระดับ (Multi-Level Analysis) เป็นวิธีการวิเคราะห์ข้อมูลสำหรับข้อมูลที่มีลักษณะเป็นเชิงลำดับ กล่าวคือเป็นการวิเคราะห์ที่มีการคำนึงถึงระดับขั้นของข้อมูล โดยการแบ่งแยกความผันแปรที่เกิดขึ้นจากแต่ละระดับทำให้สามารถแก้ปัญหาจากการวิเคราะห์ดังกล่าวได้ ในกรณีที่โครงสร้างของข้อมูลมี 2 ระดับ ตัวแบบเชิงเส้นแบบเชิงลำดับ หรือตัวแบบเชิงเส้นพหุระดับ (Hierarchical Linear Model or Multi-Level Linear Model) จะมีรูปแบบทั่วไปดังนี้

ตัวแบบระดับที่ 1 (Level-1 Model)

(1.1) $$y_{ij} = \beta_{0j} + \sum_{q=1}^{Q} \beta_{qj} x_{qij} + r_{ij}$$

ตัวแบบระดับที่ 2 (Level-2 Model)

(1.2) $$\beta_{qj} = \gamma_{q0} + \sum_{s=1}^{S} \gamma_{qs} w_{sj} + u_{qj}$$

โดยที่ $i = 1,2,3,..., n_j$, $j = 1,2,3,...,J$

จาก (1.1) และ (1.2) จะได้ตัวแบบรวม (Combined Model) ดังนี้

(1.3) $$y_{ij} = \gamma_{00} + \sum_{q=1}^{Q} \gamma_{q0} x_{qij} + \sum_{s=1}^{S} \gamma_{0s} w_{sj} + \sum_{s=1}^{S} \sum_{q=1}^{Q} \gamma_{qs} w_{sj} x_{qij} + u_{0j} + \sum_{q=1}^{Q} u_{qj} x_{qij} + r_{ij}$$

จาก (1.1) และ (1.2) ในแต่ละกลุ่ม j มีข้อมูล n_j ชุด ดังนั้นสามารถเขียนในรูปของตัวแบบทั่วไป (General Linear Model) ได้เป็นดังนี้

ตัวแบบทั่วไประดับที่ 1 (Level-1 General Model)

(1.4) $$\underline{Y}_j = X_j \underline{\beta}_j + \underline{r}_j$$

ตัวแบบทั่วไประดับที่ 2 (Level-2 General Model)

(1.5) $$\underline{\beta}_j = W_j \underline{\gamma} + \underline{u}_j$$

แทนสมการ (1.4) ลงใน (1.3) จะได้ตัวแบบรวม

(1.6) $$\underline{Y}_j = X_j W_j \underline{\gamma} + X_j \underline{u}_j + \underline{r}_j$$

โดยอาจเรียกตัวแบบดังกล่าวว่า ตัวแบบสัมประสิทธิ์การถดถอยสุ่ม (Random Coefiicients Model)

เมื่อ $\underline{Y}_j$ แทนเวกเตอร์ของตัวแปรตามของกลุ่มที่ j ขนาด $(n_j \times 1)$

X_j แทนเมตริกซ์ของตัวแปรอิสระในระดับที่ 1 ของกลุ่มที่ j ขนาด $(n_j \times (Q+1))$

$\underline{\beta}_j$ แทนเวกเตอร์ของสัมประสิทธิ์ถดถอยของกลุ่มที่ j ขนาด $((Q+1)\times 1)$

$\underline{r}_j$ แทนเวกเตอร์ของความคลาดเคลื่อนสุ่มในระดับที่ 1 (Level-1 Random Errors) ของ กลุ่มที่ j ขนาด $(n_j \times 1)$

W_j แทนเมตริกซ์ของตัวแปรอิสระในระดับที่ 2 ของกลุ่มที่ j ขนาด $((Q+1)\times F)$

$\underline{\gamma}$ แทนเวกเตอร์ของอิทธิพลคงที่ (Fixed Effects) ขนาด $(F \times 1)$

และ $\underline{u}_j$ แทนเวกเตอร์ของอิทธิพลสุ่มหรือความคลาดเคลื่อนสุ่มในระดับที่ 2 (Level-2 Random Effects or Level-2 Random Errors) ของกลุ่มที่ j ขนาด $((Q+1)\times 1)$

ข้อสมมติเบื้องต้นของตัวแบบคือ เวกเตอร์ของความคลาดเคลื่อนสุ่มในระดับที่ 1 มีการแจกแจงแบบปกติหลายตัวแปร ที่มีค่าเฉลี่ย $\underline{0}$ และเมตริกซ์ความแปรปรวนร่วม $\sigma^2 \cdot I_{(n_j)}$ หรือ $\underline{r}_j \overset{iid}{\sim} N_{n_j}(\underline{0},\sigma^2 \cdot I)$ และเวกเตอร์ความคลาดเคลื่อนสุ่มในระดับที่ 2 มีการแจกแจงแบบปกติหลายตัวแปร มีค่าเฉลี่ย $\underline{0}$ และเมตริกซ์ความแปรปรวนร่วม T หรือ $\underline{u}_j \overset{iid}{\sim} N_{Q+1}(\underline{0},T)$ ค่าคลาดเคลื่อนทั้งในระดับที่ 1 และระดับที่ 2 เป็นอิสระจากตัวแปรอิสระจากทั้งสองระดับ และเป็นอิสระซึ่งกันและกัน โดยที่เมตริกซ์ความแปรปรวนร่วม T มีสมาชิกคือ $\tau_{qq'}$ ในกรณีที่ $q = q'$ จะได้ τ_{qq} เป็นค่าความแปรปรวนของ u_q และในกรณีที่ $q \neq q'$ จะได้ $\tau_{qq'}$ เป็นค่าความแปรปรวนร่วมระหว่าง u_q กับ $u_{q'}$ และจะเรียกรวม T และ σ^2 ว่า ส่วนประกอบความแปรปรวน (Variance-Covariance Component)

พารามิเตอร์ในตัวแบบ (1.6) สามารถแบ่งออกได้เป็นสองส่วนด้วยกันคือ ส่วนของพารามิเตอร์สัมประสิทธิ์การถดถอย ($\underline{\gamma}$) หรืออาจเรียกว่า พารามิเตอร์อิทธิพลคงที่ (Fixed Effect Parameters) และส่วนของพารามิเตอร์ส่วนประกอบความแปรปรวน ซึ่งแบ่งเป็นพารามิเตอร์ความแปรปรวนในระดับที่ 1 (σ^2) และพารามิเตอร์ส่วนประกอบความแปรปรวนในระดับที่ 2 (T) หรืออาจเรียกรวมว่าพารามิเตอร์อิทธิพลสุ่ม (Random Effects Parameters) แนวทางการประมาณค่าพารามิเตอร์ในตัวแบบดังกล่าว แบ่งออกได้เป็น 2 แนวทางใหญ่คือ การประมาณค่าพารามิเตอร์โดยใช้แนวทางวิธีภาวะน่าจะเป็นสูงสุด (Maximum Likelihood Method: ML) และการใช้แนวทางของเบส์ (Bayesian Method) โดยในงานวิจัยนี้จะพิจารณาเฉพาะการประมาณ

ค่าพารามิเตอร์โดยอาศัยแนวทางวิธีภาวะน่าจะเป็นสูงสุดเท่านั้น ในกรณีเมื่อความคลาดเคลื่อนสุ่มในทั้งสองระดับมีการแจกแจงแบบปกติ ตัวประมาณที่ได้จากวิธีภาวะน่าจะเป็นสูงสุดจะเป็นตัวประมาณที่เอนเอียง (Biased) แต่มีคุณสมบัติที่ดีคือเป็นตัวประมาณที่มีคุณสมบัติไม่เอนเอียงเมื่อใกล้อนันต์ (Asymptotic Unbiased) มีความคงเส้นคงวา (Consistent) และเป็นตัวประมาณที่มีประสิทธิภาพ (Efficient)

จากการที่ตัวประมาณที่ได้จากวิธีภาวะน่าจะเป็นสูงสุดเป็นตัวประมาณที่มีความเอนเอียงโดยเฉพาะเมื่อตัวอย่างมีขนาดเล็ก จึงได้มีการพัฒนาวิธีการประมาณ เรียกว่าวิธีภาวะน่าจะเป็นสูงสุดแบบมีข้อจำกัด (Restricted Maximum Likelihood Method : REML) ซึ่งเป็นตัวประมาณที่ไม่เอนเอียง (Unbiased Estimator)

ในการหาค่าประมาณของพารามิเตอร์ด้วยวิธีภาวะน่าจะเป็นสูงสุดดังกล่าว จะใช้การประมาณโดยที่อัลกอริทึม (Algorithm) ซึ่งในตัวแบบเชิงเส้นพหุระดับมีอยู่ด้วยกันหลายวิธีเช่น วิธี Expectation-Maximization (EM) วิธี Fisher Scoring วิธี Iterative Generalized Least Square (IGLS) วิธี Restricted Iterative Generalized Least Square (RIGLS) โดยวิธีประมาณดังกล่าวเป็นวิธีประมาณที่ส่งค่าไปกลับจนสุดท้ายแล้วค่าประมาณจะลู่เข้า (converge) สู่ค่าประมาณที่เหมาะสม ซึ่งโดยทั่วไปแล้วเมื่อข้อสมมติเบื้องต้นของตัวแบบเป็นจริง ทุกอัลกอริทึมจะได้ค่าประมาณที่ใกล้เคียงกัน แต่ถ้าตัวแบบมีความซับซ้อนมากขึ้นแต่ละอัลกอริทึมจะแตกต่างกันทั้งความเร็วในการลู่เข้าสู่ค่าประมาณ ความยากง่ายในการคำนวณ และในบางกรณีบางอัลกอริทึมอาจไม่ได้ค่าประมาณที่เหมาะสมเช่นค่าประมาณของส่วนประกอบความแปรปรวนอาจมีค่าเป็นศูนย์หรือมีค่าติดลบ (Snijder and Bosker ,1999: 82-83)

จากที่กล่าวมาในข้างต้นจะพบว่าขนาดตัวอย่างเป็นปัจจัยหนึ่งที่ส่งผลต่อคุณภาพของตัวประมาณ ในทางทฤษฎีเมื่อข้อสมมติเบื้องต้นเป็นจริงวิธีภาวะน่าจะเป็นสูงสุด กับวิธีภาวะน่าจะเป็นสูงสุดแบบมีข้อจำกัด จะให้ค่าประมาณของสัมประสิทธิ์ถดถอย และค่าประมาณของความแปรปรวนในระดับที่ 1 (Level-1 Variance) ในตัวแบบที่แตกต่างกันเล็กน้อย แต่ค่าประมาณของส่วนประกอบความแปรปรวนในระดับที่ 2 (Level-2 Variance Components) จะมีความแตกต่างกันเมื่อขนาดตัวอย่างในระดับที่ 2 มีขนาดเล็ก โดยที่ค่าประมาณของส่วนประกอบความแปรปรวนในระดับที่ 2 ที่ได้จากวิธีภาวะน่าจะเป็นสูงสุดแบบมีข้อจำกัดจะให้ค่าประมาณที่มีความผิดพลาดน้อยกว่าค่าประมาณจากวิธีภาวะน่าจะเป็นสูงสุด แต่ถ้าขนาดตัวอย่างในระดับที่ 2 มีขนาดใหญ่ ค่าประมาณที่ได้จากทั้งสองวิธีก็จะมีค่าไม่แตกต่างกัน มีงานวิจัยหลายชิ้นที่ยืนยันสอดคล้องกันว่าขนาดตัวอย่างในระดับที่ 2 มีความสำคัญมากกว่าขนาดตัวอย่างในระดับที่ 1 ซึ่งในการศึกษาพบว่าเมื่อขนาดตัวอย่างในระดับที่ 2 มีขนาดเล็ก ค่าประมาณของค่าคลาดเคลื่อนมาตรฐานจะมีค่าต่ำกว่าปกติ ในปี ค.ศ. 2003 แมส และ ฮอคค์ ได้ศึกษาเกี่ยวกับผลกระทบของขนาดตัวอย่าง

ต่อคุณภาพของตัวประมาณในกรณีที่ขนาดตัวอย่างมีขนาดเล็ก โดยศึกษาเฉพาะวิธีภาวะน่าจะเป็นสูงสุดแบบมีข้อจำกัด (REML) ด้วยวิธี Restricted Iterative Generalized Least Square พบว่าค่าประมาณของอิทธิพลคงที่ (Fixed Effect Parameters) ไม่มีความเอนเอียง แต่ค่าคลาดเคลื่อนมาตรฐานของตัวประมาณพารามิเตอร์จะเริ่มประมาณค่าได้ผิดพลาดเมื่อขนาดตัวอย่างในระดับที่ 2 มีขนาดเล็กกว่า 30 หน่วย (Van Der Leeden, Busing and Meijer, 1997: 1; Snijder and Bosker, 1999: 82-83; Raudenbush and Bryk, 2002 : 51-55, 436-438; Mass and Hox, 2003: 9-10)

ปัจจัยอีกปัจจัยหนึ่งที่ส่งผลต่อคุณภาพต่อตัวประมาณคือ ค่าสัมประสิทธิ์สหสัมพันธ์ภายในกลุ่ม (Intraclass Correlation: ICC) ซึ่งเป็นค่าที่วัดระดับความสัมพันธ์ของค่าสังเกตภายในกลุ่ม คำนวณได้จากสัดส่วนของความผันแปรระหว่างกลุ่ม (Between Group Variance) ต่อความผันแปรทั้งหมด (Total Variance) โดยถ้าข้อมูลชุดใดยิ่งมีค่า ICC สูงการเลือกการใช้การวิเคราะห์พหุระดับกับข้อมูลชุดนั้นก็จะมีความเหมาะสมมากกว่าการใช้การวิเคราะห์แบบระดับเดียวมากขึ้น (Goldstein, 1995: 5; Raudenbush and Bryk, 2002: 36, 71-72)

ในทางปฏิบัติข้อมูลมักไม่ได้มีคุณสมบัติตรงตามข้อสมมติเบื้องต้นโดยเฉพาะข้อสมมติของความเป็นปกติของความคลาดเคลื่อนสุ่ม กล่าวคือความคลาดเคลื่อนสุ่มอาจมีการแจกแจงแบบไม่ปกติ (Nonnormality) โดยในตัวแบบเชิงเส้นพหุระดับนี้ การละเมิดข้อสมมติเบื้องต้นดังกล่าวจึงสามารถเกิดขึ้นได้ทั้ง 2 ระดับ ในทางทฤษฎีถ้าข้อสมมติของการแจกแจงความคลาดเคลื่อนสุ่มดังกล่าวไม่เป็นจริง ตัวประมาณด้วยวิธีภาวะน่าจะเป็นสูงสุด จะยังเป็นตัวประมาณที่มีคุณสมบัติคงเส้นคงวา และไม่เอนเอียงเมื่อใกล้อนันต์อยู่ แต่ค่าคลาดเคลื่อนมาตรฐานของตัวประมาณจะประมาณค่าได้ผิดพลาด ทำให้การประมาณค่าแบบช่วง และการทดสอบสมมติฐานผิดพลาด (Goldstein, 1995: 14-15; Raudenbush and Bryk, 2002: 266-267, 274-275)

วิธีการหนึ่งเพื่อแก้ไขปัญหาที่เกิดขึ้นจากปัญหาค่าคลาดเคลื่อนไม่ได้มีการแจกแจงแบบปกติ คือการปรับการคำนวณค่าคลาดเคลื่อนมาตรฐานให้ถูกต้อง โดยใช้ตัวประมาณค่าคลาดเคลื่อนมาตรฐานที่มีความแกร่ง (Robust Standard Error) เช่น ตัวประมาณค่าคลาดเคลื่อนมาตรฐานที่มีความแกร่งของฮูเบอร์ไวท์ (Huber/White Robust Standard Error Estimator) หรือตัวประมาณแซนวิช (Sandwich Estimator) ทำให้การประมาณค่าแบบช่วง และการทดสอบสมมติฐานมีความถูกต้องมากขึ้น (Goldstein, 1995: 14-15; Raudenbush and Bryk, 2002: 276-280)

ในงานวิจัยนี้สนใจที่จะศึกษาผลกระทบต่อค่าประมาณ และค่าคลาดเคลื่อนมาตรฐานเมื่อการแจกแจงความคลาดเคลื่อนสุ่มมีการแจกแจงแบบไม่ปกติ โดยพิจารณาในกรณีที่การแจกแจง

ความคลาดเคลื่อนสุ่มในระดับที่ 2มีความเบ้ขวา และมีความโด่งมากกว่า หรือมีความแบนราบมากกว่าการแจกแจงแบบปกติ และพิจารณาเพื่อเป็นแนวทางในการกำหนดขนาดตัวอย่างในระดับที่ 2 ที่เหมาะสมว่าควรมีขนาดเท่าใด เมื่อการแจกแจงของความคลาดเคลื่อนสุ่มมีลักษณะดังที่กล่าวในข้างต้น และสุดท้ายศึกษาว่าตัวประมาณค่าคลาดเคลื่อนมาตรฐานที่มีความแกร่งของฮูเบอร์ไวท์ สามารถนำมาใช้แก้ปัญหาการคำนวณค่าคลาดเคลื่อนมาตรฐานผิดพลาดเมื่อการแจกแจงความคลาดเคลื่อนสุ่มมีการแจกแจงแบบไม่ปกติได้หรือไม่

วัตถุประสงค์ของการวิจัย

1. เพื่อศึกษาผลกระทบต่อตัวประมาณพารามิเตอร์ และค่าคลาดเคลื่อนมาตรฐาน เมื่อการแจกแจงความคลาดเคลื่อนสุ่มมีการแจกแจงแบบไม่ปกติ จากการประมาณค่าพารามิเตอร์ 2 วิธีคือ วิธี Iterative Generalized Least Square (IGLS) และวิธี Restricted Iterative Generalized Least Square (RIGLS)
2. เพื่อศึกษาถึงขนาดตัวอย่างในระดับที่ 2 ที่เหมาะสม เมื่อการแจกแจงความคลาดเคลื่อนสุ่มมีความเบ้ขวา และมีความโด่งมากกว่าหรือแบนราบมากกว่าการแจกแจงแบบปกติ
3. เพื่อศึกษาว่าตัวประมาณค่าคลาดเคลื่อนมาตรฐานที่มีความแกร่งของฮูเบอร์ไวท์ มีความเหมาะสมในการประมาณค่าคลาดเคลื่อนมาตรฐานผิดพลาดเมื่อความคลาดเคลื่อนสุ่มมีการแจกแบบไม่ปกติ

สมมติฐานของการวิจัย

1. ผลกระทบของความคลาดเคลื่อนสุ่มในระดับที่ 2 ที่มีการแจกแจงแบบไม่ปกติไม่ได้ส่งผลกระทบต่อค่าความเอนเอียงสัมพัทธ์ของค่าประมาณพารามิเตอร์ในทั้งสองวิธี
2. วิธี Restricted Iterative Generalized Least Square จะให้ค่าประมาณของส่วนประกอบความแปรปรวนในระดับที่ 2 ที่ถูกต้องกว่าวิธี Iterative Generalized Least Square
3. ตัวประมาณค่าคลาดเคลื่อนมาตรฐานที่มีความแกร่งของฮูเบอร์ไวท์ จะให้ค่าประมาณที่ถูกต้องมากกว่า ตัวประมาณค่าคลาดเคลื่อนมาตรฐานแบบปกติ เมื่อค่าคลาดเคลื่อนสุ่มมีการแจกแจงแบบไม่ปกติ

ข้อตกลงเบื้องต้น

1. ตัวแบบที่ใช้ศึกษาเป็นตัวแบบสัมประสิทธิ์ถดถอยสุ่ม ที่มี 2 ระดับโดยที่ในระดับที่ 1 มีตัวแปรอิสระ 1 ตัวดังนี้

   ตัวแบบระดับที่ 1 (Level-1 Model)

   (1.7) $$y_{ij} = \beta_{0j} + \beta_{1j}x_{ij} + r_{ij}$$

   ตัวแบบระดับที่ 2 (Level-2 Model)

   (1.8ก) $$\beta_{0j} = \gamma_{00} + u_{0j}$$

   (1.8ข) $$\beta_{1j} = \gamma_{01} + u_{1j}$$

   ตัวแบบรวม (Combined Model)

   (1.9) $$y_{ij} = \gamma_{00} + \gamma_{01}x_{ij} + u_{0j} + u_{1j}x_{ij} + r_{ij}$$

2. ค่าคลาดเคลื่อนสุ่มในระดับที่ 1 ($\underline{r}_j$) จะกำหนดให้มีการแจกแจงแบบปกติ หลายตัวแปร (Multivariate Normal Distribution) และค่าคลาดเคลื่อนสุ่มในระดับที่ 2 ($\underline{u}_j = (u_{0j}, u_{1j})^T$) จะกำหนดให้เป็นเวกเตอร์สุ่มที่มีการแจกแจงแบบเอกโพเนนเชียลพาวเวอร์แบบไม่สมมาตรสองตัวแปร (Bivariate Exponential Power Distribution) โดยที่กำหนดให้ u_{0j} และ u_{1j} เป็นอิสระกัน $\forall j$
3. ขนาดตัวอย่างในระดับที่ 1 ที่ติดกลุ่มอยู่ในระดับที่ 2 จะกำหนดให้มีขนาดตัวอย่างเท่ากันทุกกลุ่ม กล่าวคือ $n_j = n \quad \forall j$ เมื่อ n_j คือขนาดตัวอย่างในระดับที่ 1 ในกลุ่มที่ j
4. อัลกอริทึมที่ใช้ในการศึกษาประมาณพารามิเตอร์จะใช้วิธี Iterative Generalized Least

   Square (IGLS) และ Restricted Iterative Generalized Least Square (RIGLS)
5. ตัวประมาณค่าคลาดเคลื่อนมาตรฐานคือ ตัวประมาณค่าคลาดเคลื่อนมาตรฐานแบบปกติ หรือตัวประมาณค่าคลาดเคลื่อนมาตรฐานจากตัวแบบ(Asymtotic Standard Error Estimator) และตัวประมาณค่าคลาดเคลื่อนมาตรฐานที่มีความแกร่งของฮูเบอร์ไวท์ (Huber/White Standard Error Estimator)

ขอบเขตของการวิจัย

1. ตัวแบบที่ใช้ในงานวิจัยคือ ตัวแบบเชิงเส้นเชิงลำดับชั้นอย่างง่าย หรือสัมประสิทธิ์ถดถอยแบบสุ่มอย่างง่าย กล่าวคือเป็นตัวแบบที่มีสมการใน (1.9)

2. ข้อมูลที่นำมาใช้ในงานวิจัยในครั้งนี้จะเป็นข้อมูลจากสมการ (1.9) โดยสร้าง X_{ij} เป็นค่าใดๆ ในที่นี้กำหนดให้ $X_{ij} \sim N(3,16)$ ส่วนค่า Y สร้างมาจากสมการ (1.9) ที่มีค่าความคลาดเคลื่อนแจกแจงตามกำหนด
3. ในงานวิจัยครั้งนี้จะกำหนดให้การแจกแจงของค่าคลาดเคลื่อนสุ่มในระดับที่ 1 มีการแจกแจงแบบปกติหลายตัวแปรที่มีค่าเฉลี่ย $\underline{0}$ และเมตริกซ์ความแปรปรวนร่วม $\sigma^2 \cdot I_{(n)}$ และการแจกแจงความคลาดเคลื่อนสุ่มในระดับที่ 2 มีการแจกแจงแบบเอกโพเนนเชียลพาวเวอร์แบบไม่สมมาตรสองตัวแปรที่มีค่าเฉลี่ย $\underline{0}$ และเมตริกซ์ความแปรปรวนร่วม Σ พารามิเตอร์ความโด่ง α และพารามิเตอร์ความเบ้ κ กล่าวคือ

   (1.10) $\underline{r}_j \sim N_{n_j}(\underline{0}, \sigma^2 I_{(n_j)})$ และ $\underline{u}_j = (u_{0j}, u_{1j})^T \sim EP_2(\underline{0}, \Sigma, \alpha, \kappa)$

   โดยที่กำหนดให้ u_{0j} และ u_{1j} เป็นอิสระกัน $\forall j$
4. กำหนดให้ค่าสัมประสิทธิ์สหสัมพันธ์ภายในกลุ่ม (Intra-class Correlation : ICC)[1] มีค่าเท่ากับ 0.1 0.2 และ 0.3
5. กำหนดให้พารามิเตอร์ในสมการ (1.9) ดังนี้

   5.1) พารามิเตอร์อิทธิพลคงที่ หรือสัมประสิทธิ์ความถดถอย $\gamma_{00} = 1$ และ $\gamma_{01} = 2$

   5.2) ความแปรปรวนของค่าคลาดเคลื่อนสุ่มในระดับที่ 1 คือ $\sigma^2 = 0.5$

   5.3) ส่วนประกอบความแปรปรวนของค่าคลาดเคลื่อนสุ่มในระดับที่ 2 จะกำหนดให้ $\tau_{00} = \tau_{11}$ โดยที่คำนวณจากค่าสัมประสิทธิ์สหสัมพันธ์ภายในกลุ่ม ซึ่งมีค่าเท่ากับ 0.1, 0.2 และ0.3 และ กำหนดให้ $Cov(u_{oj}, u_{1j}) = \tau_{01} = 0$ ซึ่งมีรายละเอียดดังในตารางที่ 1.1

ตารางที่ 1.1 แสดงค่าสัมประสิทธิ์สหสัมพันธ์ และส่วนประกอบความแปรปรวนในงานวิจัย

ICC[2]	σ^2	$\tau_{00} = \tau_{11}$	τ_{01}
0.1		0.0556	
0.2	0.5	0.1250	0
0.3		0.2143	

1 เนื่องจากค่าสัมประสิทธิ์สหสัมพันธ์ภายในกลุ่มคำนวณมาจากสัดส่วนระหว่างความผันแปรระหว่างกลุ่ม (Between group Variance)กับความผันแปรรวม(Total variance) กล่าวคือ $ICC = \tau_{00}/(\tau_{00} + \sigma^2)$ ซึ่งในตัวแบบสัมประสิทธิ์ถดถอยแบบสุ่มมีค่าความผันแปรรวมไม่คงที่ ดังนั้นในงานวิจัยนี้จึงทำการกำหนดค่าสัมประสิทธิ์สหสัมพันธ์ภายในกลุ่มในกรณีที่ค่า x=0

2 โดยทั่วไปแล้วในการวิจัยทางการศึกษาค่าสัมประสิทธิ์สหสัมพันธ์ภายในกลุ่มมักมีค่าต่ำกว่า 0.2 และในการวิจัยทางสังคมศาสตร์อื่นๆ มักพบว่าค่าสัมประสิทธิ์สหสัมพันธ์ภายในกลุ่มมักมีค่ามากกว่า 0.33 (Mass and Hox, 2003: 4)

6. จะสนใจศึกษาเมื่อกำหนดพารามิเตอร์ความโด่งและพารามิเตอร์ความเบ้ โดยกำหนดค่าจากสัมประสิทธิ์ความโด่ง และสัมประสิทธิ์ความเบ้ (Ayebo and Kozubowski, 2003) เป็นดังนี้

   กรณีที่ 1 กำหนดให้พารามิเตอร์ความเบ้คงที่โดยมีค่าเท่ากับ 1 ($\kappa = 1$) และพารามิเตอร์ความโด่งเป็น α =1, 1.25, 1.6, 3, 8 และ32

   กรณีที่ 2 กำหนดพารามิเตอร์ความโด่งคงที่โดยมีค่าเท่ากับ 2 ($\alpha = 2$) และพารามิเตอร์ความเบ้เป็น κ=0.8518, 0.6437 และ0.1

   โดยที่ค่าสัมประสิทธิ์ความเบ้ และความโด่งของในแต่ละสถานการณ์ต่างๆข้างต้นแสดงไว้ในตารางที่ 1.2

ตารางที่1.2 แสดงค่าพารามิเตอร์ความเบ้ พารามิเตอร์ความโด่ง สัมประสิทธิ์ความเบ้ และสัมประสิทธิ์ความโด่ง

พาราเตอร์ ความเบ้	**พารามิเตอร์ ความโด่ง**	**สัมประสิทธิ์ ความเบ้**	**สัมประสิทธิ์ ความโด่ง**	**ลักษณะการแจกแจง**
0.8518	2	0.2500	3.0449	
0.6437	2	0.6000	3.2683	เบ้ขวา
0.1	2	0.9904	3.8555	(Positive Skew)
1	1	0.0000	6.0000	
1	1.25	0.0000	4.5272	โด่งกว่าปกติ
1	1.6	0.0000	3.5527	(Leptokurtic)
1	3	0.0000	2.4184	
1	8	0.0000	1.9234	แบนราบกว่าปกติ
1	32	0.0000	1.8102	(Platykurtic)

เนื่องจากค่าของพารามิเตอร์ตำแหน่ง และพารามิเตอร์สเกลของการแจกแจงแบบเอกโพเนนเชียลพาวเวอร์แบบไม่สมมาตร ไม่ได้มีค่าเท่ากับค่าเฉลี่ย และความแปรปรวนของตัวแปรสุ่มที่มีการแจกแจงดังกล่าว ดังนั้นเพื่อเป็นการจำกัดให้ค่าเฉลี่ยของตัวแปรสุ่มที่มีการแจกแจงดังกล่าวมีค่าเท่ากับ 0 และความแปรปรวนเท่ากับที่กำหนดไว้ในแต่ละสถานการณ์ จึงมีการคำนวณค่าพารามิเตอร์ตำแหน่ง และพารามิเตอร์สเกลในแต่ละสถานการณ์ให้สอดคล้องกับค่าเฉลี่ยและความแปรปรวนที่ได้กำหนดไว้ ดังรายละเอียดในตารางที่ 1.3 และ 1.4

ตารางที่ 1.3 แสดงการกำหนดค่าพารามิเตอร์ตำแหน่งในแต่ละสถานการณ์ ที่ระดับสัมประสิทธิ์สหสัมพันธ์ 0.1, 0.2 และ 0.3

พารามิเตอร์ความเบ้	**พารามิเตอร์ความโด่ง**	**พารามิเตอร์ตำแหน่ง** ICC 0.1	0.2	0.3
0.8518	2	-0.0595	-0.0892	-0.1168
0.6437	2	-0.1501	-0.2250	-0.2946
0.1	2	-0.3078	-0.4615	-0.6043
1	1	0	0	0
1	1.25	0	0	0
1	1.6	0	0	0
1	3	0	0	0
1	8	0	0	0
1	32	0	0	0

ตารางที่ 1.4 แสดงการกำหนดค่าพารามิเตอร์สเกลในแต่ละสถานการณ์ ที่ระดับสัมประสิทธิ์สหสัมพันธ์ 0.1, 0.2 และ 0.3

พารามิเตอร์ความเบ้	**พารามิเตอร์ความโด่ง**	**พารามิเตอร์สเกล** *ICC* 0.1	0.2	0.3
0.8518	2	0.1072	0.2409	0.4130
0.6437	2	0.0855	0.1922	0.3294
0.1	2	0.0030	0.0068	0.0117
1	1	0.0278	0.0625	0.1072
1	1.25	0.0521	0.1171	0.2009
1	1.6	0.0837	0.1881	0.3224
1	3	0.1489	0.3349	0.5741
1	8	0.1767	0.3973	0.6811
1	32	0.1719	0.3864	0.6624

7. ขนาดตัวอย่างที่ใช้ทำการศึกษาแบ่งออกได้เป็น 2 ส่วน คือ
   7.1) ขนาดตัวอย่างในระดับที่ 1 หรือขนาดของกลุ่ม (level-1 sample sizes or sample group sizes) คือ 5, 10 และ15
   7.2) ขนาดตัวอย่างในระดับที่ 2 หรือจำนวนกลุ่มตัวอย่าง (level-2 sample sizes or number of sample groups) คือ 15, 30 และ50

ตารางที่ 1.5 แสดงขนาดของกลุ่ม จำนวนกลุ่มตัวอย่าง และขนาดตัวอย่างรวมในการวิจัย

ขนาดของกลุ่ม	**จำนวนกลุ่มตัวอย่าง**	**ขนาดตัวอย่างรวม**
5	15	75
5	30	150
5	50	250
10	15	150
10	30	300
10	50	500
15	15	225
15	30	450
15	50	750

8. ข้อมูลที่ใช้ในการศึกษาวิจัยครั้งนี้ใช้โปรแกรมคอมพิวเตอร์จำลองด้วยเทคนิคมอนติ-คาร์โล กระทำซ้ำ 500 รอบ ในแต่ละสถานการณ์

เกณฑ์การตัดสินใจ

1. ความเอนเอียงสัมพัทธ์ (Relative Bias : RB)

เป็นเกณฑ์ที่ใช้วัดความเอนเอียงของกลุ่มของค่าประมาณซึ่งแบ่งการวัดออกเป็น 3 กลุ่มคือ กลุ่มอิทธิพลคงที่ (γ_{00}, γ_{01}) กลุ่มความแปรปรวนในระดับที่ 1 (σ^2) และกลุ่มความแปรปรวนในระดับที่ 2 (τ_{00}, τ_{11}) โดยมีสูตรที่ใช้คำนวณดังนี้

$$RB = \{(Mean(\hat{\underline{\theta}}) - \underline{\theta})^T W^{-1} (Mean(\hat{\underline{\theta}}) - \underline{\theta})\}^{1/2}$$

เมื่อ $\underline{\theta}$ คือเวกเตอร์ของพารามิเตอร์ของกลุ่มที่ต้องการประมาณ

$Mean(\hat{\underline{\theta}})$ คือเวกเตอร์ของค่าเฉลี่ยของค่าประมาณพารามิเตอร์ของกลุ่มพารามิเตอร์ที่ต้องการศึกษาที่ได้จากการจำลอง 500 รอบ

W คือเมตริกซ์ทแยงมุมถ่วงน้ำหนัก (Diagonal Weighted Matrix) ที่มีสมาชิกในแนว ทแยงมุมเป็นค่ากำลังสองของค่าพารามิเตอร์จริงที่ทำการประมาณ

2. รากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองระหว่างค่าประมาณค่าคลาดเคลื่อนมาตรฐานกับค่าคลาดเคลื่อนมาตรฐานแบบมอนติคาร์โลสัมพัทธ์[3] (Root of Relative Mean Square Error between Estimated Standard Error and Monte Carlo Standard Error : RMSE)

เป็นเกณฑ์ที่ใช้วัดความผิดพลาดในการประมาณค่าคลาดเคลื่อนมาตรฐานจากแต่ละวิธี โดยเทียบกับค่าคลาดเคลื่อนมาตรฐานที่แท้จริง (True Standard Error) โดยมีหลักเกณฑ์การวัดดังต่อไปนี้

เนื่องจากค่าคลาดเคลื่อนมาตรฐานแบบมอนติคาร์โล จะมีค่าลู่เข้าสู่ค่าคลาดเคลื่อนมาตรฐานที่แท้จริง (True Standard Error) ถ้าจำนวนรอบในการจำลองมีมากพอ (Cheong, Fotiu and Raudenbush; 2001: 421) ดังนั้นเกณฑ์การตัดสินใจว่าวิธีการใดที่มีความถูกต้องมากที่สุด จะใช้ค่าคลาดเคลื่อนมาตรฐานแบบมอนติคาร์โล (Monte Carlo Standard Error: MCSE) เป็นค่าที่ใช้เปรียบเทียบกับค่าประมาณของค่าคลาดเคลื่อนมาตรฐานที่ได้ทั้งจากตัวแบบ และจากวิธีประมาณค่าคลาดเคลื่อนมาตรฐานที่มีความแกร่งของฮูเบอร์ไวท์ โดยมีสูตรการคำนวณดังนี้

$$RMSE = \{\frac{1}{500}\sum_{k=1}^{500}(\underline{SE_{ijk}} - \underline{MCSE_{ij}})^T W^{-1}(\underline{SE_{ijk}} - \underline{MCSE_{ij}})\}^{1/2}$$

เมื่อ $\underline{SE_{ijk}}$ แทน เวกเตอร์ของค่าคลาดเคลื่อนมาตรฐานของพารามิเตอร์กลุ่มที่ i ที่ประมาณจากวิธีประมาณที่ j รอบการจำลองที่ k

$\underline{MCSE_{ij}}$ แทนเวกเตอร์ของค่าคลาดเคลื่อนมาตรฐานแบบมอนติคาร์โลที่ประมาณจากวิธีประมาณที่ j รอบการจำลองที่ k

โดยที่ i = 1, 2 และ 3; j=1, 2, 3 และ 4; k=1, 2, …, 500

[3] โดยปรกติแล้วการเปรียบเทียบความถูกต้องในการประมาณค่าคลาดเคลื่อนมาตรฐานควรใช้เกณฑ์สัมประสิทธิ์ความเชื่อมั่น ประกอบกับเกณฑ์ความยาวช่วงความเชื่อมั่นเฉลี่ย ซึ่งในการประมาณค่าแบบช่วงของพารามิเตอร์ในตัวแบบการถดถอยเชิงลำดับชั้นนี้ สูตรการประมาณค่าแบบช่วงจะใช้การแจกแจงปรกติมาตรฐาน (Standard Normal Distribution) ซึ่งเป็นการแจกแจงโดยประมาณของตัวสถิติโดยจะประมาณค่าได้ดีเมื่อตัวอย่างมีขนาดใหญ่ แต่ในการวิจัยนี้ได้กำหนดขนาดตัวอย่างไว้ในขนาดที่ค่อนข้างเล็ก ดังนั้นการแจกแจงของตัวสถิติจะไม่ได้มีการแจกแจงแบบปรกติมาตรฐาน หรือใกล้เคียง ทำให้การประมาณค่าแบบช่วงมีความผิดพลาด ส่งผลให้เกณฑ์การตัดสินใจและข้อสรุปมีความผิดพลาดตามไปด้วย ดังนั้นเพื่อหลีกเลี่ยงความผิดพลาดที่เกิดขึ้นดังกล่าว ผู้วิจัยจึงได้ใช้เกณฑ์ดังที่ได้เสนอไว้ในข้างต้น เพื่อเปรียบเทียบความผิดพลาดในการประมาณค่าคลาดเคลื่อนมาตรฐานของแต่ละวิธี

ประโยชน์ที่คาดว่าจะได้รับ

1. เพื่อทราบถึงผลกระทบที่มีต่อตัวประมาณพารามิเตอร์ในตัวแบบ และค่าคลาดเคลื่อนมาตรฐานของตัวประมาณ เมื่อการแจกแจงความคลาดเคลื่อนสุ่มมีการแจกแจงแบบไม่ปกติ
2. เพื่อเป็นแนวทางในการเลือกขนาดตัวอย่างในระดับที่ 2 ที่เหมาะสม เมื่อการแจกแจงความคลาดเคลื่อนสุ่มมีการแจกแจงแบบไม่ปกติ
3. เพื่อทราบว่าตัวประมาณค่าคลาดเคลื่อนมาตรฐานของฮูเบอร์ไวท์ เหมาะสมที่จะใช้ในการแก้ไขปัญหาการประมาณค่าคลาดเคลื่อนมาตรฐานผิดพลาด เมื่อการแจกแจงความคลาดเคลื่อนสุ่มมีการแจกแจงแบบไม่ปกติได้หรือไม่