บทที่ ๑

บทนำ

๑.๑ ที่มาของปัญหา

การประมาณค่าแวเรียนซ์ คอมโปเนนท์ (Variance Component) ในแบบจำลองผสม (Mixed Model) เป็นที่สนใจของนักสถิติมาหลายปีจนกระทั่งปัจจุบัน ดังที่ทราบกันอยู่แล้วว่าในกรณีข้อมูลที่เก็บรวบรวมได้ในแบบจำลองเชิงสุ่ม (Random Model) มีลักษณะสมดุลย์ (Balanced Data) และผลกระทบ (effects) ทุกตัวมีการแจกแจงแบบปกติ (Normal Distribution) การประมาณค่าแวเรียนซ์ คอมโปเนนท์ จะได้จากวิธีวิเคราะห์ความแปรปรวน (Analysis of Variance Method) ซึ่งค่าประมาณที่ได้จะเป็นทั้งค่าประมาณที่ไม่เอนเอียง (Unbiased Estimator) และมีความแปรปรวนน้อยที่สุด (Minimum Variance) ในกรณีที่ข้อมูลที่เก็บรวบรวมได้มีลักษณะไม่สมดุลย์ (Unbalanced Data) การประมาณค่าแวเรียนซ์ คอมโปเนนท์ จะทำได้หลายวิธี เช่น

๑./ วิธีของ Henderson วิธีที่ ๑)

๒./ วิธีของ Henderson วิธีที่ ๒)

๓./ วิธีของ Henderson วิธีที่ ๓)

๔./ วิธีการแบบภาวะน่าจะเป็นสูงสุด (Maximum Likelihood)

๕./ วิธีการแบบภาวะน่าจะเป็นสูงสุดแบบมีข้อจำกัด (Restricted Maximum Likelihood)

๖./ วิธี MIVQUE (Minimum Variance Quadratic Unbiased Estimator)

๗./ วิธี MINQUE (Minimum Norm Quadratic Unbiased Estimator)

๘./ วิธี Iterative MINQUE

ในการวิจัยนี้จะกล่าวถึงการประมาณค่าแวเรียนซ์ คอมโปเนนท์ เมื่อข้อมูลที่เก็บรวบรวมได้มีลักษณะไม่สมดุลย์ ในการแจกแจงแบบสองทาง เฉพาะบางวิธี คือ วิธีการแบบภาวะน่าจะเป็นสูงสุด วิธีการแบบภาวะน่าจะเป็นสูงสุด แบบมีข้อจำกัด วิธี MINQUE และวิธี Iterative MINQUE

วิธีการแบบภาวะน่าจะเป็นสูงสุด ในที่นี้เรียกสั้น ๆ ว่าวิธี MLE Hartley และ Rao เป็นผู้ริเริ่มนำวิธีการนี้มาใช้ ในปี พ.ศ. ๒๕๑๐ วิธีนี้จะได้ค่าประมาณของทั้งผลกระทบคงที่ (fixed effect) และแวเรียนซ์ คอมโปเนนท์ ค่าประมาณที่ได้มีคุณสมบัติที่ดีคือ เมื่อแทนค่าประมาณในฟังก์ชั่นภาวะน่าจะเป็น (likelihood function) จะทำให้ฟังก์ชั่นมีค่าสูงสุด

ในปี พ.ศ. ๒๕๑๖ Miller ได้พิจารณาคุณสมบัติของค่าประมาณของผลกระทบคงที่และแวเรียนซ์ คอมโปเนนท์ ในแบบจำลองต่าง ๆ และพบว่าค่าประมาณนั้นเป็นค่าที่สอดคล้อง (Consistent) และมีการแจกแจงแบบปกติเมื่อข้อมูลใกล้อนันต์ (Asymptotically Normal) เมื่อการทดลองถูกทำซ้ำ ๆ หรือเมื่อขนาดของตัวประกอบเชิงสุ่ม (random factor) เพิ่มขึ้นจนเข้าใกล้ค่าอนันต์ (infinity)

อย่างไรก็ตามวิธี ML มีข้อเสียคือการหาค่าประมาณทำได้ยุ่งยาก เพราะค่าประมาณเป็นผลจากการแก้สมการที่ไม่ใช่เชิงเส้น (Nonlinear equation) และจะต้องมีข้อจำกัดเพื่อให้ค่าประมาณของแวเรียนซ์ คอมโปเนนท์ เป็นบวกเสมอ เพราะค่าจริงของแวเรียนซ์ คอมโปเนนท์ มีค่าเป็นบวกดังที่กำหนดในขอบเขตของพารามิเตอร์ (parameter space) นอกจากนี้วิธี ML ยังไม่คำนึงถึงผลของการประมาณผลกระทบคงที่ที่มีต่อขั้นแห่งความเป็นอิสระ (degree of freedom) นักสถิติหลายคนพยายามปรับปรุงวิธี ML ให้ดีขึ้นอยู่ตลอดมา ในปี พ.ศ. ๒๕๑๔ Patterson และ Thompson ได้แก้ข้อเสียของวิธีนี้โดยหาค่าประมาณของแวเรียนซ์ คอมโปเนนท์ จากส่วนหนึ่งของฟังก์ชั่น ภาวะน่าจะเป็นซึ่งไม่ขึ้นกับผลกระทบคงที่ โดยค่าประมาณของแวเรียนซ์ คอมโปเนนท์

จะทำให้ฟังก์ชั่นภาวะน่าจะเป็นซึ่งไม่ขึ้นกับผลกระทบคงที่มีค่าสูงสุด วิธีการนี้มีชื่อว่า วิธีการแบบภาวะน่าจะเป็นสูงสุดแบบมีข้อจำกัดหรือเรียกสั้น ๆ ว่า วิธี REML

วิธีการอื่น ๆ นอกเหนือจากการใช้ฟังก์ชั่นภาวะน่าจะเป็น ที่นิยมใช้คือวิธี MINQUE ซึ่งนักสถิติชื่อ C.R. Rao เป็นผู้คิดขึ้น วิธีการนี้จะให้ค่าประมาณของแวเรียนซ์ คอมโปเนนท์ เพียงอย่างเดียวในแบบจำลองผสม ซึ่งมีคุณสมบัติ minimum norm และไม่เอนเอียง อย่างไรก็ตามวิธีการ MINQUE ต้องใช้ค่าสมมติเบื้องต้น (prior value) ของแวเรียนซ์ คอมโปเนนท์

ในการประมาณแวเรียนซ์ คอมโปเนนท์ โดยวิธีนี้นั้น ถ้าค่าสมมติเบื้องต้นเป็นสัดส่วนกับค่าจริง ค่าประมาณจะมีคุณสมบัติที่ดียิ่งขึ้น คือ จะมีความแปรปรวนน้อยที่สุด ในทางปฏิบัติจะไม่ทราบว่าค่าสมมติเบื้องต้นเป็นสัดส่วนกับค่าจริงหรือไม่ เพราะไม่ทราบค่าจริงของแวเรียนซ์ คอมโปเนนท์ ได้มีผู้ปรับปรุงวิธีการ MINQUE โดยใช้ชื่อว่า Iterative MINQUE ซึ่งเป็นการใช้วิธี MINQUE หลายขั้นโดยค่าสมมติเบื้องต้นในแต่ละขั้นจะเป็นค่าประมาณของแวเรียนซ์ คอมโปเนนท์ ในขั้นก่อน โดยคาดว่าค่าประมาณจากวิธี Iterative MINQUE จะใกล้กับค่าจริงมากกว่าค่าประมาณจากวิธี MINQUE เป็นที่น่าสังเกตว่าค่าประมาณที่เป็นลบไม่ได้ที่ได้จากวิธี Iterative MINQUE จะให้ผลเหมือนกันทุกประการกับวิธี REML เพราะเป็นผลจากการแก้สมการเดียวกัน และใช้ค่าสมมติเบื้องต้นในแต่ละขั้นเป็นค่าประมาณของแวเรียนซ์ คอมโปเนนท์ ในขั้นก่อนเหมือนกัน

ดังนั้น คุณสมบัติของค่าประมาณจากวิธี Iterative MINQUE เมื่อเป็นลบไม่ได้จะเหมือนกันทุกประการกับคุณสมบัติของค่าประมาณจากวิธี REML

การประมาณค่าแวเรียนซ์ คอมโปเนนท์ โดยวิธี ML ทำได้ไม่ง่ายนัก โดยค่าประมาณจะได้จากการแก้สมการ

$$(x'v^{-1}x)\beta = x'v^{-1}y \qquad \ldots\ldots\ldots\ldots (1.1)$$

$$tr(v^{-1}u_i u_i') = (y-x\beta)'v^{-1}u_i u_i' v^{-1}(y-x\beta)$$

$$i = 1, 2, \ldots, p \qquad \ldots\ldots\ldots\ldots (1.2)$$

โดยวิธี successive approximation ซึ่งต้องกำหนดค่าสมมติของแวเรียนซ์คอมโปเนนท์ เบื้องต้นในสมการ (1.1) แล้วแก้สมการหา $\hat{\beta}$ แทน $\hat{\beta}$ ในสมการ (1.2) แล้วแก้สมการหาค่าประมาณของแวเรียนซ์ คอมโปเนนท์ แล้วจึงเริ่มรอบใหม่ โดยแทนค่าประมาณของแวเรียนซ์ คอมโปเนนท์ ในสมการ (1.1) ในกรณีที่ค่าประมาณของแวเรียนซ์ คอมโปเนนท์ มีค่าเป็นลบจะเริ่มวิธี ML ใหม่จากแบบจำลองที่ไม่มีผลกระทบเชิงสุ่มของค่าประมาณที่เป็นลบ การเริ่มใหม่นี้จะใช้ค่าสมมติเบื้องต้นเป็นค่าที่ได้จากค่าประมาณในรอบก่อน โปรแกรมคอมพิวเตอร์สำหรับหาค่าประมาณของแวเรียนซ์คอมโปเนนท์ โดยวิธี ML แสดงไว้ในภาคผนวก (Appendix)

ส่วนการประมาณค่าแวเรียนซ์ คอมโปเนนท์ โดยวิธี MINQUE เริ่มจากการแทนค่าสมมติเบื้องต้นของ แวเรียนซ์ คอมโปเนนท์ ในสมการ

$$S\hat{r} = q$$

โดยที่ S คือ p x p เมทริกส์ ซึ่ง ij^{th} สมาชิก คือ

$$tr(QV_i QV_j), \quad Q = v^{-1} - v^{-1}x\,(x'v^{-1}x)^{-1}\,x'v^{-1}$$

$$V = \sum_{i=1}^{p} \sigma_i^2 V_i, \quad V_i = u_i u_i'$$

$\hat{r}$ เป็นเวคเตอร์ขนาด p x 1 ซึ่ง i^{th} สมาชิก คือ $\hat{\sigma}_i^2$

และ q เป็นเวคเตอร์ขนาด p x 1 ซึ่ง i^{th} สมาชิก คือ $Y'QV_iQY$

แล้วแก้สมการ $\sum_{j=1}^{p} tr(QV_iQV_j)\,\hat{\sigma}_j^2 = Y'QV_iQY$

ค่าประมาณที่ได้อาจจะมีค่าเป็นลบ สำหรับวิธี Iterative MINQUE เป็นการทำต่อจากวิธี MINQUE โดยการนำค่าประมาณที่ได้กลับไปเป็นค่าสมมติเบื้องต้นในรอบต่อไป จนกว่าค่าจะไม่เปลี่ยนแปลง ซึ่งค่าประมาณที่ได้แยกเป็น ๒ กรณีคือ ในกรณีที่ค่าประมาณมีค่าเป็นลบได้กับในกรณีที่กำหนดให้ค่าประมาณมีค่าเป็นบวกเสมอ โดยพิจารณาว่าค่าประมาณที่ได้มีค่าเป็นลบหรือไม่ ถ้าเป็นลบให้ค่าประมาณนั้นมีค่าเป็นศูนย์ แล้วเริ่มวิธี MINQUE ใหม่จากแบบจำลองที่ไม่มีผลกระทบเชิงสุ่มที่มีค่าประมาณเป็นลบดังกล่าว ถ้าเป็นบวกก็จะใช้ค่าประมาณที่ได้เป็นค่าสมมติเบื้องต้นในขั้นต่อไป โปรแกรมคอมพิวเตอร์สำหรับหาค่าประมาณของแวเรียนซ์ คอมโปเนนท์ โดยวิธี MINQUE แสดงในภาคผนวกเช่นกัน

๑.๒ วัตถุประสงค์ของการวิจัย

วัตถุประสงค์ที่สำคัญของการวิจัยมีดังนี้

๑. ศึกษาวิธีการประมาณค่าแวเรียนซ์ คอมโปเนนท์ ในแบบจำลองผสมของการแจกแจงแบบสองทาง และข้อมูลที่เก็บรวบรวมได้มีลักษณะไม่สมดุลย์ โดย

๑.๑ วิธี MINQUE

๑.๒ วิธี Iterative MINQUE ซึ่งกำหนดว่าค่าประมาณเป็นลบได้

๑.๓ วิธี Iterative MINQUE ซึ่งกำหนดว่าค่าประมาณเป็นลบไม่ได้ หรือวิธี REML

๑.๔ วิธี ML

๒. สร้างโปรแกรมคอมพิวเตอร์สำหรับหาค่าประมาณของแวเรียนซ์ คอมโปเนนท์ สำหรับวิธีในข้อ ๑

๓. ศึกษาคุณสมบัติของค่าประมาณที่ได้จากทั้ง ๔ วิธี

๔. เปรียบเทียบค่าประมาณและคุณสมบัติของค่าประมาณที่ได้จากทั้ง ๔ วิธี

๕. เสนอแนะวิธีการประมาณค่าแวเรียนซ์ คอมโปเนนท์ ในแบบจำลองผสมของการแจกแจงแบบอื่น เมื่อข้อมูลที่เก็บรวบรวมได้มีลักษณะไม่สมดุลย์

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

๑.๓ สมมติฐานของการทำวิจัย

การประมาณค่าแวเรียนซ์ คอมโปเนนท์ โดยวิธี Iterative MINQUE ซึ่งกำหนดให้ค่าประมาณมีค่าเป็นบวกเสมอ จะให้ค่าประมาณที่มีคุณสมบัติที่ดีกว่าการประมาณค่าแวเรียนซ์ คอมโปเนนท์ โดยวิธีอื่น ๆ เพราะค่าประมาณดังกล่าวมีคุณสมบัติที่ดีหลายประการคือ

๑. ค่าประมาณไม่เอนเอียง

๒. ค่าประมาณมีความแปรปรวนน้อยที่สุด ถ้าค่าสมมติเบื้องต้นของแวเรียนซ์ คอมโปเนนท์ ในขั้นสุดท้ายเป็นสัดส่วนกับค่าจริงของแวเรียนซ์ คอมโปเนนท์

๓. เป็น translation invariant

๔. หาค่าประมาณได้ไม่ยาก

๕. เป็นค่าที่สอดคล้อง

๖. มีการกระจายแบบปกติเมื่อข้อมูลใกล้อนันต์

๑.๔ ความสำคัญหรือประโยชน์ที่คาดว่าจะได้รับจากการทำวิจัยนี้

๑. การวิจัยนี้จะเป็นแนวทางให้นักวิจัยเข้าใจวิธีประมาณค่าแวเรียนซ์ คอมโปเนนท์ ในแบบจำลองเชิงเส้น

๒. จากโปรแกรมคอมพิวเตอร์ สำหรับหาค่าประมาณของแวเรียนซ์ คอมโปเนนท์ โดยวิธีต่าง ๆ ซึ่งแสดงไว้ในภาคผนวก ทำให้นักวิจัยหาค่าประมาณได้สดวกขึ้น

๓. จากผลการวิจัยนี้จะเป็นแนวทางให้นักวิจัยเลือกวิธีประมาณค่าแวเรียนซ์ คอมโปเนนท์ ในแบบจำลองผสมของการแจกแจงแบบสองทาง เมื่อข้อมูลที่เก็บรวบรวมได้มีลักษณะไม่สมดุลย์ โดยพิจารณาค่าประมาณที่ได้และคุณสมบัติของค่าประมาณจากแต่ละวิธี