การเปรียบเทียบเกณฑ์การคัดเลือกจำนวนปัจจัยในการวิเคราะห์ปัจจัยเชิงสถิติ

นางสาวพรรณนิภา รินทระ

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาสถิติศาสตรมหาบัณฑิต
สาขาวิชาสถิติ ภาควิชาสถิติ
คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2552

A COMPARISON OF CRITERIA FOR DETERMINING NUMBER OF FACTORS IN STATISTICAL FACTOR ANALYSIS

Miss Pannipa Rintara

A Thesis Submitted in Partial Fulfillment of the Requirements

for the Degree of Master of Science Program in Statistics

Department of Statistics

Faculty of Commerce and Accountancy

Chulalongkorn University

Academic Year 2009

หัวข้อวิทยานิพนธ์	การเปรียบเทียบเกณฑ์การคัดเลือกจำนวนปัจจัยในการวิเคราะห์ปัจจัยเชิงสถิติ
โดย	นางสาวพรรณนิภา รินทระ
สาขาวิชา	สถิติ
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	ผู้ช่วยศาสตราจารย์ ดร.เสกสรร เกียรติสุไพบูลย์

คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

..... คณบดีคณะพาณิชยศาสตร์และการบัญชี
(รองศาสตราจารย์ ดร.อรรณพ ตันละมัย)

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการ
(รองศาสตราจารย์ ดร.ธีระพร วีระถาวร)

..... อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก
(ผู้ช่วยศาสตราจารย์ ดร.เสกสรร เกียรติสุไพบูลย์)

..... กรรมการ
(รองศาสตราจารย์ ดร.กัลยา วานิชย์บัญชา)

..... กรรมการ
(รองศาสตราจารย์ ดร.สุพล ดุรงค์วัฒนา)

..... กรรมการภายนอกมหาวิทยาลัย
(อาจารย์ ดร.อัครินทร์ ไพบูลย์พานิช)

พรรณนิภา รินทระ : การเปรียบเทียบเกณฑ์การคัดเลือกจำนวนปัจจัยในการวิเคราะห์ปัจจัยเชิงสถิติ. (A COMPARISON OF CRITERIA FOR DETERMINING NUMBER OF FACTORS IN STATISTICAL FACTOR ANALYSIS) อ.ที่ปรึกษาวิทยานิพนธ์หลัก : ผศ.ดร.เสกสรร เกียรติสุไพบูลย์., 96 หน้า.

การศึกษาวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพเกณฑ์การคัดเลือกจำนวนปัจจัยในการวิเคราะห์ปัจจัย ระหว่างเกณฑ์ 10-fold Likelihood Cross-Validation (LCV) เกณฑ์การคัดเลือกจำนวนปัจจัยโดยใช้ข้อสนเทศของอากาอิเคะ (AIC) เกณฑ์การคัดเลือกจำนวนปัจจัยโดยใช้ข้อสนเทศของชวาร์ซ (SIC) และเกณฑ์การคัดเลือกจำนวนปัจจัยโดยใช้ข้อสนเทศของแฮนแนนและควินน์ (HQ) โดยใช้อัตราความถูกต้อง (%) เป็นเกณฑ์ในการเปรียบเทียบประสิทธิภาพ และทำการศึกษากับข้อมูลที่ได้จากการจำลองให้ข้อมูลมีการแจกแจงแบบปกติหลายตัวแปรที่มีเวกเตอร์ค่าเฉลี่ย $\underset{\sim}{0}$ และค่าแปรปรวนเท่ากับ 1 ซึ่งเมทริกซ์สหสัมพันธ์ได้จากการสุ่มแบบสม่ำเสมอบนเซตของเมทริกซ์สหสัมพันธ์ที่เป็นไปได้ทั้งหมด โดยการศึกษาครอบคลุมกรณีที่จำนวนตัวแปร (p) เท่ากับ 10, 20, 30 และ 40 จำนวนปัจจัยเท่ากับ 1, 2, ..., (p/2) และมีขนาดตัวอย่างเท่ากับ 200, 300, 500 และ 1,000 ซึ่งผลการวิจัยสามารถสรุปได้ดังนี้

1. กรณีจำนวนตัวแปรเท่ากับ 10 ทั้ง 4 เกณฑ์ มีประสิทธิภาพไม่แตกต่างกันเมื่อจำนวนปัจจัยไม่เกินร้อยละ 20 ของจำนวนตัวแปร และเมื่อจำนวนปัจจัยเพิ่มขึ้นมากกว่าร้อยละ 20 ของจำนวนตัวแปร โดยส่วนใหญ่เกณฑ์ SIC เป็นเกณฑ์ที่ดีที่สุด รองลงมาคือ เกณฑ์ LCV และเกณฑ์ HQ ซึ่งมีประสิทธิภาพไม่แตกต่างกัน และเกณฑ์ AIC มีประสิทธิภาพน้อยที่สุด

2. กรณีจำนวนตัวแปรเท่ากับ 20, 30 และ 40 ทั้ง 4 เกณฑ์ มีประสิทธิภาพไม่แตกต่างกันเมื่อจำนวนปัจจัยไม่เกินร้อยละ 22.5 ของจำนวนตัวแปร เมื่อจำนวนปัจจัยมากกว่าร้อยละ 22.5 แต่ไม่เกินร้อยละ 35.83 ของจำนวนตัวแปร โดยส่วนใหญ่เกณฑ์ SIC เป็นเกณฑ์ที่ดีที่สุด รองลงมาคือ เกณฑ์ HQ เกณฑ์ LCV และเกณฑ์ AIC ตามลำดับ และเมื่อจำนวนปัจจัยมากกว่าร้อยละ 35.83 ของจำนวนตัวแปร โดยส่วนใหญ่เกณฑ์ SIC เป็นเกณฑ์ที่ดีที่สุด รองลงมาคือ เกณฑ์ LCV เกณฑ์ HQ และเกณฑ์ AIC ตามลำดับ

ภาควิชา สถิติ ลายมือชื่อนิสิต พรรณนิภา

สาขาวิชา สถิติ ลายมือชื่ออ.ที่ปรึกษาวิทยานิพนธ์หลัก

ปีการศึกษา 2552

5081846326 : MAJOR STATISTICS

KEYWORDS : FACTOR ANALYSIS / CROSS-VALIDATION / INFORMATION CRITERIA

PANNIPA RINTARA : A COMPARISON OF CRITERIA FOR DETERMINING NUMBER OF FACTORS IN STATISTICAL FACTOR ANALYSIS. THESIS ADVISOR : ASSIST.PROF.SEKSAN KIATSUPAIBUL,Ph.D., 96 pp.

The purpose of this research is to compare the performance among different criteria for determining the number of factors in the statistical factor analysis. The criteria under this study includes 10-fold Likelihood Cross-Validation (LCV), Akaike's Information Criteria (AIC), Schwarz's Information Criteria (SIC) and Hannan and Quinn's Information Criteria (HQ). We adopt the percentage of the accuracy as the performance measure. The data are generated from the multivariate normal distribution with mean vector $\underset{\sim}{0}$ and variance 1. The correlation matrix is sampled from the uniform distribution on all possible correlation matrices. The number of variables (p) are 10, 20, 30 and 40. The number of factors are 1, 2, .., (p/2). The sample size are 200, 300, 500 and 1,000. The conclusions are as follows:

1. Case of 10 variables: If the number of factors is less than or equal to 20 percent of the number of variables, the performances of all criteria are the same. If the number of factors is greater than 20 percent of the number of variables, SIC is the best criteria followed by LCV, which are not different from HQ, and AIC is the worst criteria.

2. Cases of 20, 30 and 40 variables: If the number of factors is less than or equal to 22.5 percent of the number of variables, the performances of all criteria are the same. If the number of factors is greater than 22.5 percent of the number of variables but less than or equal to 35.83 percent of the number of variables, SIC is the best criteria followed by HQ, LCV and AIC, respectively. If the number of factors is greater than 35.85 percent of the number of variables, SIC is the best criteria followed by the LCV, the HQ and the AIC, respectively.

Department : Statistics Student's Signature

Field of Study : Statistics Advisor's Signature

Academic Year : 2009

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงได้ด้วยดีเนื่องจากความกรุณาของผู้ช่วยศาสตราจารย์ ดร.เสกสรร เกียรติสุไพบูลย์ อาจารย์ที่ปรึกษาวิทยานิพนธ์ ที่ได้ให้คำปรึกษาและช่วยแนะนำแก้ไขข้อบกพร่องต่างๆ ด้วยดีเสมอมา ผู้วิจัยจึงขอกราบขอบพระคุณเป็นอย่างสูงไว้ ณ โอกาสนี้

ผู้วิจัยขอกราบขอบพระคุณ รองศาสตราจารย์ ดร.ธีระพร วีระถาวร รองศาสตราจารย์ ดร.กัลยา วานิชย์บัญชา รองศาสตราจารย์ ดร.สุพล ดุรงค์วัฒนา และอาจารย์ ดร.อัครินทร์ ไพบูลย์พานิช ในฐานะประธานและกรรมการสอบวิทยานิพนธ์ ตามลำดับ ที่กรุณาตรวจสอบและแก้ไขให้วิทยานิพนธ์ฉบับนี้มีความสมบูรณ์มากยิ่งขึ้น รวมทั้งขอกราบขอบพระคุณคณาจารย์ทุกท่านที่ได้ประสิทธิประสาทวิชาความรู้ให้แก่ผู้วิจัย

สุดท้ายนี้ ผู้วิจัยใคร่ขอกราบขอบพระคุณ บิดา มารดา และครอบครัวที่ได้ให้การสนับสนุนด้านการศึกษาและคอยเป็นกำลังให้ผู้วิจัยมาโดยตลอด

สารบัญ

หน้า

สารบัญตาราง

ตารางที่ หน้า

สารบัญภาพ

บทที่ 1

บทนำ

ความเป็นมาและความสำคัญของปัญหา

ปัจจุบันการศึกษาวิจัยในด้านต่างๆ ต้องอาศัยระเบียบวิธีการทางสถิติเข้ามาช่วยในการวิเคราะห์ข้อมูลหลายตัวแปร ซึ่งเมทริกซ์ความแปรปรวนร่วม (Covariance Matrix) ก็เป็นส่วนหนึ่งที่สำคัญยิ่งในการวิเคราะห์ข้อมูลหลายตัวแปร ดังนั้นการประมาณเมทริกซ์ความแปรปรวนร่วมด้วยวิธีปกติเมื่อมีจำนวนข้อมูลไม่เพียงพอต่อการประมาณ ทำให้ค่าประมาณที่ได้ไม่ดีพอและก่อให้เกิดข้อผิดพลาดในการวิเคราะห์ข้อมูลหลายตัวแปรทางสถิติอื่นๆ ที่เกี่ยวข้อง ดังนั้นเพื่อเป็นการแก้ปัญหาดังกล่าวจึงนำการวิเคราะห์ปัจจัย (Factor Analysis) มาช่วยในการประมาณ แต่ก่อนที่จะทำการประมาณได้นั้น การกำหนดจำนวนปัจจัยให้เหมาะสมในการวิเคราะห์ปัจจัยก็เป็นเรื่องสำคัญอย่างยิ่ง

การวิเคราะห์ปัจจัย เป็นวิธีการทางสถิติวิธีการหนึ่งที่มีเป้าหมาย คือ การแสดงค่าแปรปรวนและค่าแปรปรวนร่วมของตัวแปร อีกทั้งยังเป็นวิธีที่ช่วยลดจำนวนตัวแปรอีกวิธีการหนึ่งด้วย โดยไม่มีการแบ่งข้อมูลออกเป็นตัวแปรอิสระและตัวแปรตาม ซึ่งวิธีการนี้สามารถลดจำนวนตัวแปรได้ โดยการศึกษาโครงสร้างความสัมพันธ์ของตัวแปรเดิมและสร้างตัวแปรใหม่ที่เรียกว่า ปัจจัย (Factor) ขึ้นมา ซึ่งตัวแปรที่มีความสัมพันธ์กันจะอยู่ในปัจจัยเดียวกันและตัวแปรที่มีความสัมพันธ์กันน้อยหรือไม่มีความสัมพันธ์กันจะอยู่กันคนละปัจจัย ดังนั้นปัจจัยที่สร้างขึ้นมาจะไม่มีความสัมพันธ์กันและยังสื่อความหมายถึงตัวแปรเดิมที่อยู่ในปัจจัยเดียวกันด้วย เช่น ข้อมูลคะแนนสอบของนักเรียน ซึ่งประกอบด้วยวิชา ภาษาฝรั่งเศส ภาษาอังกฤษ เลขคณิต และพีชคณิต เมื่อวิเคราะห์ข้อมูลดังกล่าวด้วยการวิเคราะห์ปัจจัยพบว่า สามารถสร้างปัจจัยใหม่ได้ 2 ปัจจัย คือ ปัจจัยที่ 1 คะแนนความสามารถทางด้านคณิตศาสตร์ ประกอบด้วยวิชาเลขคณิต และพีชคณิต ปัจจัยที่ 2 คะแนนความสามารถทางด้านภาษา ประกอบด้วยภาษาฝรั่งเศสและภาษาอังกฤษ เป็นต้น

ในหลายปีที่ผ่านมามีนักวิจัยหลายท่าน ได้ให้ความสนใจศึกษาเกี่ยวกับการวิเคราะห์ปัจจัยและได้นำวิธีการวิเคราะห์ปัจจัยไปประยุกต์ใช้กับงานด้านอื่นๆ นอกเหนือจากงานด้านสถิติ ไม่ว่าจะเป็นงานด้านการแพทย์ ด้านการเงิน หรือด้านวิทยาศาสตร์ เป็นต้น ซึ่งผู้วิจัยเองได้ให้ความสนใจในงานวิจัยของ Delores A. Conway และ Marc R. Reinganum [1] ในปี 1988 ซึ่งเป็นการศึกษาเกี่ยวกับการกำหนดจำนวนปัจจัยในตัวแบบปัจจัย สำหรับข้อมูลผลตอบแทนของหุ้น (Stock Return) ด้วยครอสวาลิเดชั่น (Cross-Validation) โดยใช้ข้อมูลจริง พบว่า

การทำครอสวาลิเดชั่นจะกำหนดจำนวนปัจจัยน้อยกว่าตัวสถิติทดสอบอัตราส่วนภาวะน่าจะเป็น (Likelihood Ratio Test Statistic) คือ สำหรับหุ้น 30-60 ตัว การทำครอสวาลิเดชั่นจะกำหนดจำนวนปัจจัย 1 ปัจจัย แต่ตัวสถิติทดสอบอัตราส่วนภาวะน่าจะเป็นจะกำหนด 4-6 ปัจจัย

ต่อมาได้มีงานวิจัยอื่นๆ ที่ได้มีการอ้างอิงงานวิจัยของ Delores A. Conway และ Marc R. Reinganum [1] อีกหลายงาน หนึ่งในงานวิจัยที่นำไปใช้อ้างอิง คือ งานวิจัยของ Michele Costa [2] ในปี 1996 เรื่อง Factor Analysis and Information Criteria ที่สนใจศึกษาเกี่ยวกับจำนวนปัจจัยที่แท้จริงในตัวแบบปัจจัยเชิงสำรวจ โดยศึกษาผ่านเกณฑ์การเปรียบเทียบทั้งหมด 10 เกณฑ์ ได้แก่ ตัวสถิติทดสอบอัตราส่วนภาวะน่าจะเป็น (Likelihood Ratio Test : LR) กระบวนการของครอสวาลิเดชั่น (Cross-Validation) เกณฑ์การคัดเลือกจำนวนปัจจัยโดยใช้ข้อสนเทศของอากาอิเกะ (Akaike's Information Criteria : AIC) เกณฑ์การคัดเลือกจำนวนปัจจัยโดยใช้ข้อสนเทศของอากาอิเกะที่ทำการแปลงเทอมที่สองเป็น αh (Akaike's Information Criteria: AIC_α) โดยที่ α = 3 และ 4 เกณฑ์การคัดเลือกจำนวนปัจจัยโดยใช้ข้อสนเทศของชวาร์ซ (Schwarz's Information Criteria : SIC) และเกณฑ์การคัดเลือกจำนวนปัจจัยโดยใช้ข้อสนเทศของแฮนแนนและควินน์ (Hannan and Quinn's Information Criteria : HQ_c)โดยเทอมที่สองคือ 2hcln(ln(n)) เมื่อ c = 1, 2, 3 และ 4 โดยใช้ข้อมูลที่ได้จากการจำลองเมื่อกำหนดจำนวนปัจจัยที่แท้จริงมาก่อน ซึ่งข้อมูลที่ได้จากการจำลองนั้นมาจากตัวแบบปัจจัย (Factor Model) คือ

$$\underset{n\times p}{X} = \underset{n\times m}{F}\ \underset{m\times p}{L} + \underset{n\times p}{\varepsilon}$$

หรือ

$$\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix} = \begin{pmatrix} f_{11} & f_{12} & \cdots & f_{1m} \\ f_{21} & f_{22} & \cdots & f_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ f_{n1} & f_{n2} & \cdots & f_{nm} \end{pmatrix} \begin{pmatrix} \ell_{11} & \ell_{12} & \cdots & \ell_{1p} \\ \ell_{21} & \ell_{22} & \cdots & \ell_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \ell_{m1} & \ell_{m2} & \cdots & \ell_{mp} \end{pmatrix} + \begin{pmatrix} e_{11} & e_{12} & \cdots & e_{1p} \\ e_{21} & e_{22} & \cdots & e_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ e_{n1} & e_{n2} & \cdots & e_{np} \end{pmatrix}$$

โดยที่ n คือ ขนาดตัวอย่าง

p คือ จำนวนตัวแปร

m คือ จำนวนปัจจัย

F คือ เมทริกซ์ขนาด nxm ของคะแนนปัจจัย (Factor Score) ที่มีการแจกแจงแบบปกติหลายตัวแปร (Multivariate Normal Distribution) มีเมทริกซ์ความแปรปรวนร่วมเท่ากับเมทริกซ์เอกลักษณ์ (Identity Matrix) ($Cov(F)=I_{mxm}$)

ε คือ เมทริกซ์ขนาด nxp ของค่าเฉพาะ ที่มีการแจกแจงแบบปกติหลายตัวแปร และมีเมทริกซ์ความแปรปรวนร่วมเท่ากับเมทริกซ์เอกลักษณ์ ($Cov(\varepsilon)=I_{pxp}$) เช่นกัน

L คือ เมทริกซ์ขนาด mxp ของ Factor Loading ที่ได้จากการวิเคราะห์ปัจจัยของ p asset return ที่ถูกสุ่มมาจาก 100 asset returns ของตลาดหุ้นมิลานระหว่างปี 1986 ถึง 1989

จากงานวิจัยข้างต้นของ Michele Costa [2] พบว่า ข้อมูลที่นำมาทำการวิเคราะห์เป็นข้อมูลที่จำลองมาจากข้อมูลจริงเพียงชุดเดียว คือ 100 asset returns ของตลาดหุ้นมิลานระหว่างปี 1986 ถึง 1989 ดังนั้น ข้อมูลจำลองที่ได้มาจึงยังไม่ครอบคลุม เนื่องจาก Factor Loading ได้มาจากข้อมูลเพียงชุดเดียวและค่าเฉพาะถูกจำกัดให้มีเมทริกซ์ความแปรปรวนร่วมเท่ากับเมทริกซ์เอกลักษณ์เท่านั้น ทำให้ผู้วิจัยสนใจศึกษาผลการเปรียบเทียบเกณฑ์การคัดเลือกจำนวนปัจจัยที่ดีที่สุด 4 เกณฑ์ ของ Michele Costa [2] ได้แก่

1. (10-fold) Likelihood Cross-Validation (LCV)
2. เกณฑ์การคัดเลือกจำนวนปัจจัยโดยใช้ข้อสนเทศของอากาอิเกะ (AIC)
3. เกณฑ์การคัดเลือกจำนวนปัจจัยโดยใช้ข้อสนเทศของชวาร์ซ (SIC)
4. เกณฑ์การคัดเลือกจำนวนปัจจัยโดยใช้ข้อสนเทศของแฮนแนนและควินน์(HQ)

โดยปรับปรุงข้อมูลที่นำมาวิเคราะห์ให้มีความครอบคลุมมากยิ่งขึ้นโดยใช้วิธี Onion ซึ่งเสนอโดย Ghosh และ Henderson [3] ซึ่งเป็นวิธีการจำลองเมทริกซ์สหสัมพันธ์ขึ้นมา ซึ่งจะกล่าวถึงรายละเอียดในบทที่ 2 เอกสารและงานวิจัยที่เกี่ยวข้อง เพื่อสร้าง Factor Loading ทุกรูปแบบที่เป็นไปได้และค่าเฉพาะจะแปรไปตามส่วนที่เหลือจากตัวแบบปัจจัยของตัวอย่างที่สุ่มมาซึ่งจะมีค่ามากหรือน้อยก็ได้

โดยวิธีการสกัดปัจจัยที่ใช้ในครั้งนี้ คือ วิธีตัวประกอบหลัก (Principal Component Method) สำหรับการเปรียบเทียบประสิทธิภาพของทั้ง 4 เกณฑ์จะพิจารณาจากอัตราความถูกต้องของการคัดเลือกจำนวนปัจจัย

วัตถุประสงค์ของการวิจัย

เพื่อทำการเปรียบเทียบประสิทธิภาพเกณฑ์การคัดเลือกจำนวนปัจจัยในการวิเคราะห์ปัจจัย โดยเกณฑ์ที่ใช้ในการเปรียบเทียบ คือ

1. (10-fold) Likelihood Cross-Validation (LCV)
2. เกณฑ์การคัดเลือกจำนวนปัจจัยโดยใช้ข้อสนเทศของอากาอิเกะ (AIC)
3. เกณฑ์การคัดเลือกจำนวนปัจจัยโดยใช้ข้อสนเทศของชวาร์ซ (SIC)

4. เกณฑ์การคัดเลือกจำนวนปัจจัยโดยใช้ข้อสนเทศของแฮนแนนและควินน์ (HQ)

ขอบเขตของการวิจัย

การดำเนินงานวิจัยในครั้งนี้ มีขอบเขตการวิจัย คือ

1. ทำการศึกษาเกณฑ์การคัดเลือกจำนวนปัจจัยในการวิเคราะห์ปัจจัย 4 เกณฑ์ กับข้อมูลที่ได้จากการจำลอง โดยเกณฑ์ทั้ง 4 เกณฑ์ ได้แก่

1.1 (10-fold) Likelihood Cross-Validation (LCV)

1.2 เกณฑ์การคัดเลือกจำนวนปัจจัยโดยใช้ข้อสนเทศของอากาอิเกะ (AIC)

1.3 เกณฑ์การคัดเลือกจำนวนปัจจัยโดยใช้ข้อสนเทศของชวาร์ซ (SIC)

1.4 เกณฑ์การคัดเลือกจำนวนปัจจัยโดยใช้ข้อสนเทศของแฮนแนนและควินน์ (HQ)

2. สถานการณ์ต่างๆ ที่สนใจศึกษาจำแนกตามจำนวนตัวแปร (p) จำนวนปัจจัย (m) และขนาดตัวอย่าง (n) ได้ดังนี้

2.1 กำหนดให้จำนวนตัวแปรเท่ากับ 10 ตัวแปร จำนวนปัจจัยเท่ากับ 1,2,…,5 ปัจจัย และขนาดตัวอย่างเท่ากับ 200, 300, 500 และ 1,000

2.2 กำหนดให้จำนวนตัวแปรเท่ากับ 20 ตัวแปร จำนวนปัจจัยเท่ากับ 1,2,…,10 ปัจจัย และขนาดตัวอย่างเท่ากับ 300, 500 และ 1,000

2.3 กำหนดให้จำนวนตัวแปรเท่ากับ 30 ตัวแปร จำนวนปัจจัยเท่ากับ 1,2,…,15 ปัจจัย และขนาดตัวอย่างเท่ากับ 500 และ 1,000

2.4 กำหนดให้จำนวนตัวแปรเท่ากับ 40 ตัวแปร จำนวนปัจจัยเท่ากับ 1,2,…,20 ปัจจัย และขนาดตัวอย่างเท่ากับ 1,000

3. การเปรียบเทียบประสิทธิภาพของเกณฑ์การคัดเลือกจำนวนปัจจัยในการวิเคราะห์ปัจจัย สามารถวัดได้ด้วยอัตราความถูกต้องและยืนยันการเปรียบเทียบประสิทธิภาพด้วยการทดสอบสมมติฐานเปรียบเทียบค่าสัดส่วนความถูกต้องของการคัดเลือกจำนวนปัจจัย

ข้อตกลงเบื้องต้น

ข้อตกลงเบื้องต้นสำหรับการวิจัยในครั้งนี้ คือ

1. ข้อมูลที่นำมาทำการวิเคราะห์ปัจจัยเป็นข้อมูลที่ได้จากการจำลอง (Simulation)

2. ข้อมูลที่นำมาทำการวิเคราะห์ปัจจัยเป็นข้อมูลที่มีการแจกแจงแบบปกติหลายตัวแปร (Multivariate Normal Distribution) ที่มีเวกเตอร์ค่าเฉลี่ยเท่ากับ $\underset{\sim}{0}$ และเมทริกซ์ความแปรปรวนร่วม Σ ขนาด pxp ที่มีค่าแปรปรวนเท่ากับ 1 นั่นคือ

$$\Sigma = \begin{pmatrix} 1 & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & 1 & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & 1 \end{pmatrix}$$

3. วิธีที่ใช้ในการสกัดปัจจัยในการวิเคราะห์ปัจจัยครั้งนี้ คือ วิธีตัวประกอบหลัก (Principal Component Method)

คำจำกัดความที่ใช้ในการวิจัย

อัตราความถูกต้อง (%) = $\frac{\text{จำนวนรอบที่คัดเลือกจำนวนปัจจัยถูกต้อง}}{\text{จำนวนรอบที่ทำการวิเคราะห์ปัจจัยทั้งหมด}} \times 100$

สัดส่วนของจำนวนปัจจัยต่อจำนวนตัวแปร (%) = $\frac{\text{จำนวนปัจจัย (m)}}{\text{จำนวนตัวแปร (p)}} \times 100$

ประโยชน์ที่คาดว่าจะได้รับ

1. เพื่อเป็นแนวทางในการตัดสินใจว่า ควรใช้เกณฑ์การคัดเลือกจำนวนปัจจัยเกณฑ์ใดในการวิเคราะห์ปัจจัย ภายใต้สถานการณ์ต่างๆ จึงจะเหมาะสม

2. เพื่อเป็นแนวทางในการศึกษาเปรียบเทียบ หรือ พัฒนาเกณฑ์การคัดเลือกจำนวนปัจจัยในการวิเคราะห์ปัจจัย ในสถานการณ์อื่นๆ ต่อไป

3. เพื่อเป็นแนวทางในการศึกษาเปรียบเทียบเกณฑ์ทางสถิติอื่นๆ ที่เกี่ยวข้องต่อไป

วิธีดำเนินการวิจัย

ขั้นตอนในการดำเนินการวิจัย มีดังนี้

1. จำลองข้อมูลให้เป็นไปตามขอบเขตที่กำหนด คือ กำหนดจำนวนตัวแปร จำ - นวนปัจจัย และขนาดตัวอย่างตามขอบเขตการวิจัย โดยข้อมูลตัวอย่างแต่ละชุดที่ได้จะมีการแจกแจงแบบปกติหลายตัวแปรตามที่กำหนดไว้ข้างต้น

2. คัดเลือกจำนวนปัจจัยในการวิเคราะห์ปัจจัยตามเกณฑ์การคัดเลือกจำนวนปัจจัยแต่ละเกณฑ์

3. ตรวจสอบความถูกต้องของการคัดเลือกจำนวนปัจจัยและประเมินประสิทธิภาพของแต่ละเกณฑ์ โดยพิจารณาจากอัตราความถูกต้อง กราฟระหว่างสัดส่วนของจำนวนปัจจัยต่อจำนวนตัวแปร (%) และอัตราความถูกต้องของเกณฑ์ทั้ง 4 เกณฑ์ และทดสอบสมมติฐานเปรียบเทียบค่าสัดส่วนความถูกต้องเพื่อยืนยันผลที่ได้

4. สรุปผลที่ได้จากการวิจัย

ลำดับขั้นตอนในการเสนอผลการวิจัย

เนื่องในโอกาสเฉลิมฉลองวาระมงคลพิเศษแห่งการสถาปนามหาวิทยาลัยครบ 50 ปี บัณฑิตวิทยาลัยมหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ ได้รับเกียรติให้เป็นเจ้าภาพจัดการประชุมเสนอผลงานวิจัยระดับบัณฑิตศึกษาแห่งชาติ ครั้งที่ 14 โดยมีลำดับขั้นตอนในการเสนอผลการวิจัย ดังต่อไปนี้

1. เตรียมบทความที่ประกอบด้วยบทคัดย่อ บทนำ วัตถุประสงค์ วิธีการวิจัย ผลการวิจัย บทวิจารณ์ บทสรุป กิตติกรรมประกาศ และเอกสารอ้างอิง โดยมีรูปแบบตามที่บัณฑิตวิทยาลัยมหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือกำหนด

2. ส่งบทความฉบับวิจัยเต็มพร้อมบทคัดย่อ จำนวน 3 ชุด ตามรายละเอียดดังนี้

2.1 บทคัดย่อ 1 ชุด

2.2 บทความฉบับเต็ม (มีชื่อผู้วิจัย) 1 ชุด

2.3 บทความฉบับเต็ม (ไม่มีชื่อผู้วิจัย) 1 ชุด

3. แก้ไขบทความตามที่ผู้ทรงคุณวุฒิแนะนำเมื่อผลงานวิจัยผ่านการพิจารณาคัดเลือกผลงานให้นำเสนอแบบบรรยาย

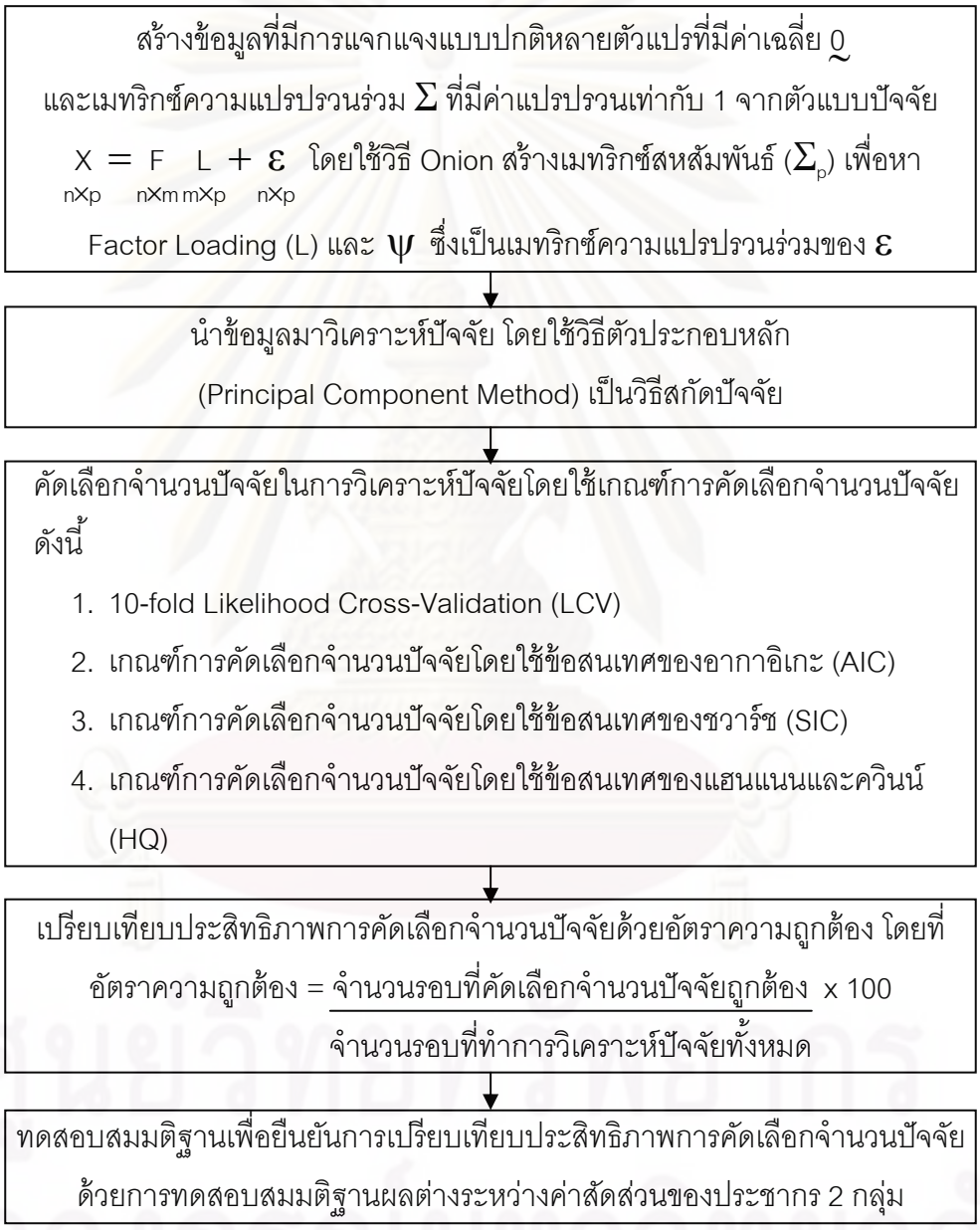
4. ส่งบทความฉบับสมบูรณ์ที่ทำการแก้ไขเรียบร้อยแล้ว

5. ผู้วิจัยนำเสนอผลงานวิจัยด้วยตนเองในวันที่ 11 กันยายน พ.ศ. 2552 เวลา 09.40 น.-10.00 น. ณ หอประชุมเบญจรัตน์ อาคารนวมินทรราชินี

บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

การศึกษาวิจัยในครั้งนี้เป็นการศึกษาการเปรียบเทียบเกณฑ์การคัดเลือกจำนวนปัจจัยในการวิเคราะห์ปัจจัยเชิงสถิติ โดยมีลำดับขั้นตอนการทดลองดังแผนภาพที่ 2.1



รูปที่ 2.1 แสดงลำดับขั้นตอนการทดลองสำหรับการคัดเลือกจำนวนปัจจัยในการวิเคราะห์ปัจจัยด้วยเกณฑ์การคัดเลือกจำนวนปัจจัยทั้ง 4 เกณฑ์

ดังนั้นแนวคิดและทฤษฎีที่เกี่ยวข้องในการศึกษาวิจัยครั้งนี้จะครอบคลุมเนื้อหาสำคัญโดยเรียงลำดับแนวคิดและทฤษฎีตามลำดับขั้นตอนการทดลองข้างต้น ดังนี้

1. วิธีการสุ่มเมทริกซ์ที่มีการแจกแจงแบบสม่ำเสมอบนเซตของเมทริกซ์สหสัมพันธ์มิติ p (Σ_p)

2. การวิเคราะห์ปัจจัย (Factor Analysis)

3. วิธีสกัดปัจจัย คือ วิธีตัวประกอบหลัก (Principal Component Method)

4. (10-fold) Likelihood Cross-Validation (LCV)

5. เกณฑ์ข้อสนเทศ (Information Criteria : IC)

5.1 เกณฑ์การคัดเลือกจำนวนปัจจัยโดยใช้ข้อสนเทศของอากาอิเกะ (Akaike's Information Criteria : AIC)

5.2 เกณฑ์การคัดเลือกจำนวนปัจจัยโดยใช้ข้อสนเทศของชวาร์ช (Schwarz's Information Criteria : SIC)

5.3 เกณฑ์การคัดเลือกจำนวนปัจจัยโดยใช้ข้อสนเทศของแฮนแนนและควินน์ (Hannan and Quinn's Information Criteria : HQ)

6. การทดสอบสมมติฐานผลต่างระหว่างค่าสัดส่วนของประชากร 2 กลุ่ม

สำหรับเนื้อหาสำคัญดังกล่าวข้างต้นจะกล่าวถึงรายละเอียดในส่วนต่อไปนี้

แนวคิดและทฤษฎี

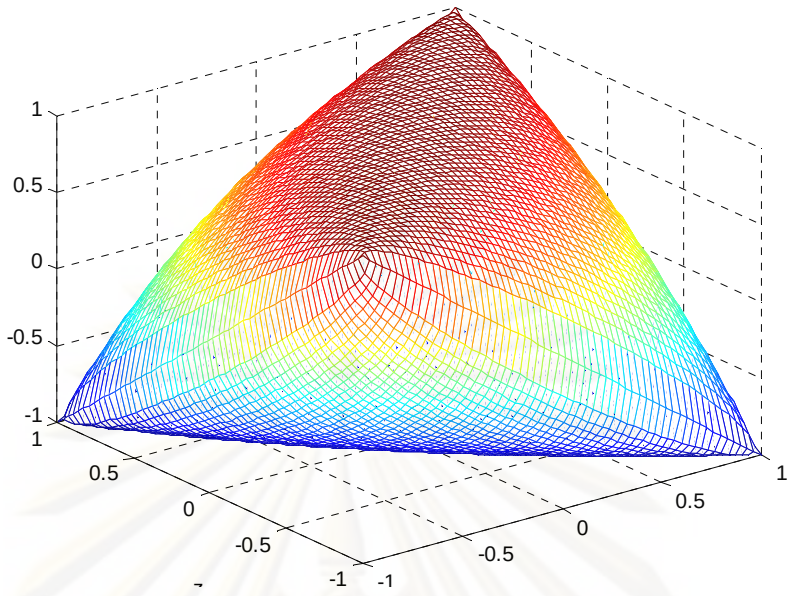
1. วิธีการสุ่มเมทริกซ์ที่มีการแจกแจงแบบสม่ำเสมอบนเซตของเมทริกซ์สหสัมพันธ์มิติ p (Σ_p)

สามารถสร้างได้ด้วยวิธี Onion [3],[4] โดยมีวัตถุประสงค์เพื่อสร้างเมทริกซ์สหสัมพันธ์ที่มีการแจกแจงแบบสม่ำเสมอบน $\Omega_p = \{\Sigma : \Sigma = \Sigma', \Sigma \succeq 0, \Sigma_{ii} = 1; i = 1, 2, \ldots, p\}$ ที่มีความเที่ยงตรงและรวดเร็ว โดย Ω_p มีคุณสมบัติดังนี้

1. เป็นเซตของเมทริกซ์สหสัมพันธ์ที่เป็นเมทริกซ์สมมาตรและกึ่งบวก (Symmetric Positive Semidefinite Matrix) ที่มีค่าบนเส้นทแยงมุมเท่ากับ 1

2. เป็นเซตย่อย (Subset) ของปริภูมิเวกเตอร์ $R^{p(p-1)/2}$

3. มีรูปร่างนูน (Convex) และมีขอบเขต (Boundary) อย่างเช่นกรณี 3 มิติ แสดงดังรูปต่อไปนี้



รูปที่ 2.2 แสดงรูปร่างของเมทริกซ์สหสัมพันธ์ใน 3 มิติ

สำหรับเมทริกซ์สุ่ม Σ ให้ Σ_k แทนเมทริกซ์ย่อยบน-ซ้าย ขนาด kxk ของเมทริกซ์ Σ และ f_k แทนความหนาแน่นส่วนริมของ Σ_k เมื่อ Σ มีฟังก์ชันความหนาแน่นร่วม $f(\Sigma) \propto 1, \quad \forall \Sigma \in \Omega_p$ และให้ q คือ เวกเตอร์ ตัวอย่างเช่น

$$\Sigma_k = \begin{pmatrix} \Sigma_{k-1} & q \\ q & 1 \end{pmatrix}$$

เรียก q ว่า ส่วนเติมเต็ม (Completion) ของ Σ_{k-1} ใน Σ_k

วิธี Onion เป็นวิธีการทำซ้ำ นั่นคือ เริ่มจากเมทริกซ์ 1 มิติ แล้วเพิ่มมิติของเมทริกซ์ขึ้นเรื่อยๆ โดยการเพิ่มเวกเตอร์ (q) ที่มีการแจกแจงที่เหมาะสม ตามแถวและหลักของเมทริกซ์ จนได้เมทริกซ์ที่มีมิติตามที่ต้องการ โดยขั้นตอนในการสร้างเมทริกซ์สหสัมพันธ์ด้วยวิธี Onion มีดังนี้

1. ให้ $\Sigma_1 = 1$ เป็นเมทริกซ์มิติ 1x1
2. สำหรับ k = 2, 3, … , p

2.1 จำลองตัวแปรสุ่ม $Y \sim \text{Beta}\left(\alpha_1 = \frac{k-1}{2}, \quad \alpha_2 = \frac{(p-k)}{2} + 1\right)$ มา 1 ตัว

2.2 ให้ $r = \sqrt{y}$

2.3 จำลองตัวแปรสุ่ม Z ที่มีการแจกแจงแบบปกติมาตรฐานที่เป็นอิสระกัน k-1 ตัว

2.4 ให้ $\theta = \frac{1}{|Z|} Z$ โดยที่ $|Z|$ คือขนาดของ Z

2.5 ให้ $w = r\theta$

2.6 ให้ $q = \Sigma_{k-1}^{1/2} w$

โดยที่ $\Sigma_{k-1}^{1/2}$ คือ ผกผัน (Inverse) ของ $\Sigma_{k-1}^{-1/2}$ ซึ่ง $\Sigma_{k-1}^{-1/2}$ คือ Upper Triangular Cholesky Factor ของ Σ_{k-1}^{-1} ($\Sigma_{k-1}^{1/2} = (\text{Cholesky Factor}(\Sigma_{k-1}^{-1}))^{-1}$) หรือ

ให้ $q = (\Sigma_{k-1}^{1/2})^t w$

โดยที่ $\Sigma_{k-1}^{1/2}$ คือ Upper Triangular Cholesky Factor ของ Σ_{k-1}

2.7 ให้ $\Sigma_k = \begin{pmatrix} \Sigma_{k-1} & q \\ q^t & 1 \end{pmatrix}$

2.8 ถ้า $k < p$ ให้ k = k+1 แล้วกลับไปทำขั้นตอน 2.1

ถ้า $k = p$ ให้ $\Sigma = \Sigma_p$

<u>หมายเหตุ</u> ขั้นตอนที่ 2.3 - 2.4 เป็นการจำลอง Unit ball ใน R^{k-1}

2. การวิเคราะห์ปัจจัย (Factor Analysis)

เป็นเทคนิคการวิเคราะห์ข้อมูลหลายตัวแปรเทคนิคหนึ่ง โดยมีจุดประสงค์เพื่ออธิบายความสัมพันธ์ระหว่างตัวแปรหลายๆ ตัวแปรด้วยตัวแปรใหม่ที่เรียกว่าปัจจัย (Factor) เพียงไม่กี่ตัว ซึ่งปัจจัยดังกล่าวไม่สามารถเก็บค่าสังเกตได้ โดยการวิเคราะห์ปัจจัยจะเป็นการจัดกลุ่มตัวแปรโดยพิจารณาความสัมพันธ์ระหว่างตัวแปร นั่นคือ ตัวแปรที่มีความสัมพันธ์กันมากจะถูกจัดให้อยู่ในปัจจัยเดียวกัน แต่ตัวแปรที่มีความสัมพันธ์กันน้อยจะถูกจัดให้อยู่คนละปัจจัย โดยข้อมูลที่นำมาวิเคราะห์ปัจจัยจะไม่มีการแบ่งตัวแปรออกเป็นตัวแปรอิสระและตัวแปรตาม ซึ่งการวิเคราะห์ปัจจัยนั้นมักจะถูกนำไปใช้ในการลดจำนวนตัวแปร แล้วนำปัจจัยที่ได้ไปใช้ในการวิเคราะห์ทางสถิติอื่นๆ ต่อไป [5]

ตัวแบบปัจจัยเชิงตั้งฉาก (The Orthogonal Factor Model)

กำหนดให้ $\underset{\sim}{X} = (X_1, X_2, \ldots, X_p)^t$ เป็นเวกเตอร์สุ่มของค่าสังเกตสำหรับ p ตัวแปร ที่มีเวกเตอร์ค่าเฉลี่ย $\underset{\sim}{\mu}$ และเมทริกซ์ความแปรปรวนร่วม Σ ซึ่งข้อสมมติของตัวแบบปัจจัย คือ ตัวแปร X สามารถเขียนให้อยู่ในรูปฟังก์ชันเชิงเส้นของตัวแปรสุ่ม $F_1, F_2, \ldots, F_m$ เรียกว่าปัจจัยร่วม (Common Factor) โดยที่ m คือ จำนวนปัจจัย และ $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_p$ เรียกว่า ค่าเฉพาะ (Error หรือ Specific Factor หรือ Unique Factor) ที่ไม่สามารถเก็บค่าสังเกตได้ ($m << p$) โดยมีสมการดังนี้ [6]

$$
\begin{aligned}
X_1 - \mu_1 &= \ell_{11}F_1 + \ell_{12}F_2 + \ldots + \ell_{1m}F_m + \varepsilon_1 \\
X_2 - \mu_2 &= \ell_{21}F_1 + \ell_{22}F_2 + \ldots + \ell_{2m}F_m + \varepsilon_2 \\
&\vdots \qquad\qquad\qquad \vdots \\
X_p - \mu_p &= \ell_{p1}F_1 + \ell_{p2}F_2 + \ldots + \ell_{pm}F_m + \varepsilon_p
\end{aligned} \tag{1}
$$

หรือเขียนอยู่ในรูปของเมทริกซ์และเวกเตอร์ได้ดังนี้

$$
\begin{pmatrix} X_1 - \mu_1 \\ X_2 - \mu_2 \\ \vdots \\ X_p - \mu_p \end{pmatrix} = \begin{pmatrix} \ell_{11} & \ell_{12} & \cdots & \ell_{1m} \\ \ell_{21} & \ell_{22} & \cdots & \ell_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \ell_{p1} & \ell_{p2} & \cdots & \ell_{pm} \end{pmatrix} \begin{pmatrix} F_1 \\ F_2 \\ \vdots \\ F_m \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_m \end{pmatrix} \tag{2}
$$

หรือ

$$
\underset{(px1)}{X - \mu} = \underset{(pxm)}{L} \quad \underset{(mx1)}{F} + \underset{(px1)}{\varepsilon} \tag{3}
$$

โดยที่ ℓ_{ij} คือ Factor Loading ของตัวแปรตัวที่ i บนปัจจัยที่ j ดังนั้น เมทริกซ์ L คือ เมทริกซ์ของ Factor Loading

ε_i คือ ค่าเฉพาะของตัวแปรที่ i (X_i)

ในการวิเคราะห์ปัจจัยนั้นได้กำหนดข้อสมมติต่างๆ เกี่ยวกับปัจจัยร่วมและค่าเฉพาะไว้ดังนี้

$$
E(F) = \underset{(m \times 1)}{0}, \quad Cov(F) = E(FF^t) = \underset{(m \times m)}{I}
$$

$$
E(\varepsilon) = \underset{(p \times 1)}{0}, \quad Cov(\varepsilon) = E(\varepsilon\varepsilon^t) = \underset{(p \times p)}{\Psi} = \begin{pmatrix} \psi_1 & 0 & \cdots & 0 \\ 0 & \psi_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \psi_p \end{pmatrix}
$$

$$
Cov(\varepsilon, F) = E(\varepsilon F^t) = \underset{(p \times m)}{0}
$$

สำหรับเป้าหมายในการวิเคราะห์ปัจจัย คือ การแสดงค่าแปรปรวนและค่าแปรปรวนร่วมของตัวแปรเดิม $X_1, X_2, \ldots, X_p$ ในรูปของ Factor Loading จำนวน pm ค่า และค่าแปรปรวนของค่าเฉพาะ p ค่า โดยเมทริกซ์ความแปรปรวนร่วมของ X คือ [5]

$$\begin{aligned}\Sigma &= \text{Cov}(X)\\ &= \text{Cov}(X-\mu)\\ &= E[(X-\mu)(X-\mu)^t]\\ &= E[(LF+\varepsilon)(LF+\varepsilon)^t]\\ &= E[(LF+\varepsilon)((LF)^t+\varepsilon^t)]\\ &= E[LF(LF)^t+\varepsilon(LF)^t+LF\varepsilon^t+\varepsilon\varepsilon^t]\\ &= LE(FF^t)L^t+E(\varepsilon F^t)L^t+LE(F\varepsilon^t)+E(\varepsilon\varepsilon^t)\\ &= LL'+\psi \end{aligned} \quad (4)$$

และความแปรปรวนร่วมระหว่าง X กับปัจจัยร่วม F ในรูปของ Factor Loading คือ

$$\begin{aligned}\text{Cov}(X,F) &= \text{Cov}(X-\mu,F)\\ &= \text{Cov}(LF+\varepsilon)\\ &= E[(LF+\varepsilon)F^t]\\ &= E[LFF^t+\varepsilon F^t]\\ &= LE(FF^t)+E(\varepsilon F^t)\\ &= L \end{aligned} \quad (5)$$

สำหรับค่าแปรปรวนของตัวแปรที่ i ที่มี m ปัจจัย จะประกอบไปด้วยส่วนสำคัญ 2 ส่วน คือ ส่วนของค่าความร่วมกัน (Communality) และค่าแปรปรวนของค่าเฉพาะ (Uniqueness หรือ Specific Variance) ดังนี้

$$\underbrace{\sigma_{ii}}_{\text{Var}(X_i)} = \underbrace{\ell_{i1}^2+\ell_{i2}^2+\ldots+\ell_{im}^2}_{\text{Communality}} + \underbrace{\psi_i}_{\text{Uniqueness}} \quad (6)$$

หรือ $h_i^2 = \ell_{i1}^2 + \ell_{i2}^2 + \ldots + \ell_{im}^2$

และ $\sigma_{ii} = h_i^2 + \psi_i \, , i = 1, 2, \ldots, p$

นั่นคือ ค่าความร่วมกันที่ i เท่ากับผลรวมของค่ายกกำลังสองของ Factor Loading ของตัวแปรที่ i บนปัจจัย m ปัจจัย ซึ่งค่าความร่วมกันเป็นค่าที่แสดงว่าตัวแปรที่ i มีส่วนร่วมในปัจจัยร่วมมากหรือน้อย คือ ถ้าค่าความร่วมกันของตัวแปรใดมีค่ามาก แสดงว่าตัวแปรนั้นมีส่วนร่วมในปัจจัยร่วมนั้นมาก หรือปัจจัยร่วมนั้นสามารถเป็นตัวแทนของตัวแปรนั้นได้

3. วิธีสกัดปัจจัย

การสกัดปัจจัยเป็นการสร้างหรือหาปัจจัยร่วมจำนวนหนึ่งซึ่งมีจำนวนน้อยกว่าจำนวนตัวแปร (m<<p) โดยให้ปัจจัยร่วมสามารถเป็นตัวแทนของตัวแปรเดิมหรือสามารถสกัดความผันแปรต่างๆ ของตัวแปรเดิมไว้ในปัจจัยร่วมได้ ซึ่งวิธีสกัดปัจจัยมีหลายวิธี เช่น วิธีตัวประกอบหลัก (Principal Component Method) วิธีแกนหลัก (Principal Axis Method) วิธีความเป็นไปได้สูงสุด (Maximum Likelihood Method) วิธีกำลังสองน้อยสุดทั่วไป (Generalized Least Square Method) เป็นต้น โดยการเลือกวิธีสกัดปัจจัยให้เหมาะสม ขึ้นกับหลักเกณฑ์ของแต่ละวิธี สำหรับผู้ที่สนใจศึกษาเกี่ยวกับวิธีสกัดปัจจัยในการวิเคราะห์ปัจจัยที่กล่าวมาข้างต้นและอีกหลายวิธีนอกจากที่กล่าวมาสามารถศึกษาเพิ่มเติมได้จากหนังสือที่เกี่ยวกับการวิเคราะห์ข้อมูลหลายตัวแปรได้ แต่ในการศึกษาวิจัยครั้งนี้ได้ใช้วิธีตัวประกอบหลักเป็นวิธีสกัดปัจจัยซึ่งจะกล่าวถึงรายละเอียดในส่วนต่อไปนี้ [5]

วิธีตัวประกอบหลัก (Principal Component Method)

เป็นวิธีที่ใช้หลักการของการวิเคราะห์ตัวประกอบหลัก (Principal Component Analysis) โดยกำหนดเวกเตอร์สุ่ม $\underset{\sim}{X}$ เป็นเวกเตอร์ค่าสังเกต ที่มีเมทริกซ์ความแปรปรวนร่วมตัวอย่าง (Sample Covariance Matrix) $\hat{\Sigma}$ ซึ่งเป็นตัวประมาณของ Σ และสามารถหาค่าประมาณของ L หรือค่า $\hat{L}$ ได้จากสมการ (4) โดยใช้ $\hat{\Sigma}$ แทน Σ ดังนี้ [7]

$$\hat{\Sigma} = \hat{L}\hat{L}^t + \hat{\psi} \qquad (7)$$

พิจารณาการแยกส่วนเมทริกซ์ $\hat{\Sigma}$ (Spectral Decomposition) คือ

$$\hat{\Sigma} = CDC^t \qquad (8)$$

เมื่อ C คือ เมทริกซ์เชิงตั้งฉาก (Orthogonal Matrix) ที่ได้จากเวกเตอร์ไอเก็นที่ปรับปกติ $(c_i^t c_i = 1)$ ของคอลัมน์ของ $\hat{\Sigma}$ และ

$$D=\begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_p \end{pmatrix}$$

โดยที่ $\lambda_1, \lambda_2, \ldots, \lambda_p$ เป็นค่าไอเก็นของ $\hat{\Sigma}$

เมื่อ $\hat{\Sigma}$ เป็นเมทริกซ์ที่เป็นบวกแน่นอน (Positive Definite Matrix) ค่าไอเก็น $\lambda_1, \lambda_2, \ldots, \lambda_p$ จึงเป็นบวกทั้งหมด ดังนั้นจึงแยกเมทริกซ์ D ได้เป็น $D=D^{1/2}D^{1/2}$

นั่นคือ

$$D^{1/2}=\begin{pmatrix} \sqrt{\lambda_1} & 0 & \ldots & 0 \\ 0 & \sqrt{\lambda_2} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \sqrt{\lambda_p} \end{pmatrix}$$

ดังนั้น

$$\hat{\Sigma}=CDC^t$$

$$=CD^{1/2}D^{1/2}C^t$$

$$=(CD^{1/2})(CD^{1/2})^t \qquad (9)$$

ซึ่งถูกจัดให้อยู่ในรูปแบบของ $\hat{\Sigma}=\hat{L}\hat{L}^t$ แล้ว แต่ $\hat{L}$ ไม่เท่ากับ $CD^{1/2}$ เนื่องจาก $CD^{1/2}$ เป็นเทริกซ์ที่มีขนาด p x p แต่ $\hat{L}$ มีขนาด p x m โดยที่ m << p ดังนั้นจึงกำหนดให้

D_1 คือ เมทริกซ์ทแยงมุมขนาด m x m ที่ประกอบด้วยค่าไอเก็นที่มีค่ามากที่สุด m ค่า คือ $\lambda_1 > \lambda_2 > \ldots > \lambda_m$

C_1 คือ เมทริกซ์ขนาด p x m ที่ประกอบด้วยเวกเตอร์ไอเก็น $c_1, c_2, \ldots, c_m$

ดังนั้นจึงสามารถประมาณ L ได้ด้วย

$$\hat{L}=C_1D_1^{1/2} \qquad (10)$$

สำหรับการประมาณเมทริกซ์ ψ หาได้จากค่าบนเส้นทแยงมุมของเมทริกซ์ $\hat{\Sigma}-\hat{L}\hat{L}^t$

$$\hat{\psi}=\begin{pmatrix} \hat{\psi}_1 & 0 & \dots & 0 \\ 0 & \hat{\psi}_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \hat{\psi}_p \end{pmatrix} \text{ โดยที่ } \hat{\psi}_i = \hat{\sigma}_{ii} - \sum_{j=1}^{m}\hat{\ell}_{ij}^2 \qquad (11)$$

การสกัดปัจจัยด้วยวิธีตัวประกอบหลักนี้พบว่า ผลรวมของค่ายกกำลังสองของแถวและหลักของ $\hat{L}$ เท่ากับค่าความร่วมกันและค่าไอเก็น ตามลำดับ ดังนั้น ค่าความร่วมกันที่ i ประมาณได้ ดังนี้

$$\hat{h}_i^2 = \sum_{j=1}^{m}\hat{\ell}_{ij}^2 \quad ; i=1,2,\dots,p \qquad (12)$$

ซึ่งเป็นผลรวมของค่ายกกำลังสองของแถวที่ i แต่ผลรวมของค่ายกกำลังสองของหลักที่ j ของ $\hat{L}$ เท่ากับค่าไอเก็นที่ j ของ $\hat{\Sigma}$ (λ_j)

$$\sum_{i=1}^{p}\hat{\ell}_{ij}^2 = \sum_{i=1}^{p}(\sqrt{\lambda_j}c_{ij})^2$$

$$= \lambda_j \sum_{i=1}^{p} c_{ij}^2$$

$$= \lambda_j \qquad (13)$$

ซึ่ง C เป็นเมทริกซ์เชิงตั้งฉาก (Orthogonal Matrix) ที่ได้จากเวกเตอร์ไอเก็นที่ปรับปกติของคอลัมน์ของ $\hat{\Sigma}$ ดังนั้นจึงมีความยาวเท่ากับ 1

หากพบว่า ข้อมูลที่มีอยู่ไม่สมเหตุสมผลที่จะใช้เมทริกซ์ความแปรปรวนร่วม สามารถปรับข้อมูลให้อยู่ในรูปมาตรฐานได้ แล้วใช้เมทริกซ์สหสัมพันธ์ (Correlation Matrix : R) แทนเมทริกซ์ความแปรปรวนร่วม

สำหรับการหาจำนวนปัจจัยร่วมในการวิเคราะห์ปัจจัย ซึ่งเป็นตัวแทนของตัวแปรต่างๆ โดยปัจจัยร่วมดังกล่าวจะใช้อธิบายความสัมพันธ์ระหว่างตัวแปรหลายๆ ตัว การพิจารณาว่า ควรมีปัจจัยร่วมกี่ปัจจัยนั้น สามารถพิจารณาได้จากหลายเกณฑ์ด้วยกัน ได้แก่

1. จำนวนปัจจัยร่วมเท่ากับจำนวนปัจจัยที่มีผลรวมของค่าแปรปรวนสะสมมากกว่าหรือเท่ากับ 80% ของค่าแปรปรวนทั้งหมด

2. จำนวนปัจจัยร่วมเท่ากับจำนวนปัจจัยที่มีคาไอเก็นมากกว่าค่าเฉลี่ยของค่าไอเก็นทั้งหมด หากใช้ $\hat{\Sigma}$ ค่าเฉลี่ยของค่าไอเก็นเท่ากับ $\frac{\sum_{i=1}^{p}\lambda_i}{p}$ หรือใช้ R ค่าเฉลี่ยค่าไอเก็นเท่ากับ 1

3. กราฟ Scree Plot ของค่าไอเก็นของ $\hat{\Sigma}$ หรือ R คือ ถ้ากราฟลดลงอย่างรวดเร็วแล้วกลายเป็นเส้นตรงที่มีความชันน้อยมาก จำนวนปัจจัยร่วมเท่ากับจำนวนปัจจัยที่มีค่าไอเก็นก่อนที่กราฟจะกลายเป็นเส้นตรง

4. การทดสอบสมมติฐานว่า m เป็นจำนวนปัจจัยร่วมที่เหมาะสม นั่นคือ

$$H_0 : \Sigma = LL^t + \Psi$$

$$H_1 : \Sigma \neq LL^t + \Psi$$

ตัวสถิติทดสอบ คือ $\chi^2 = \left(n - \frac{2p+4m+11}{6}\right)\ln\left(\frac{\hat{L}\hat{L}^t + \hat{\psi}}{|S|}\right)$

โดยที่จะปฏิเสธ H_0 ถ้า $\chi^2 > \chi^2_v$; $v = \frac{1}{2}\left[(p-m)^2 - p - m\right]$ และประมาณ $\hat{L}$ และ $\hat{\psi}$ ด้วยวิธีความเป็นไปได้สูงสุด

แต่สำหรับการวิจัยในครั้งนี้ได้เสนอเกณฑ์การคัดเลือกจำนวนปัจจัยที่สนใจทั้งหมด 4 เกณฑ์ ได้แก่

1. (10-fold) Likelihood Cross-Validation (LCV)

2. เกณฑ์การคัดเลือกจำนวนปัจจัยโดยใช้ข้อสนเทศของอากาอิเกะ (Akaike's Information Criteria : AIC)

3. เกณฑ์การคัดเลือกจำนวนปัจจัยโดยใช้ข้อสนเทศของชวาร์ช (Schwarz's Information Criteria : SIC)

4. เกณฑ์การคัดเลือกจำนวนปัจจัยโดยใช้ข้อสนเทศของแฮนแนนและควินน์ (Hannan and Quinn's Information Criteria : HQ)

ซึ่งจะกล่าวถึงรายละเอียดในหัวข้อถัดไป ดังนี้

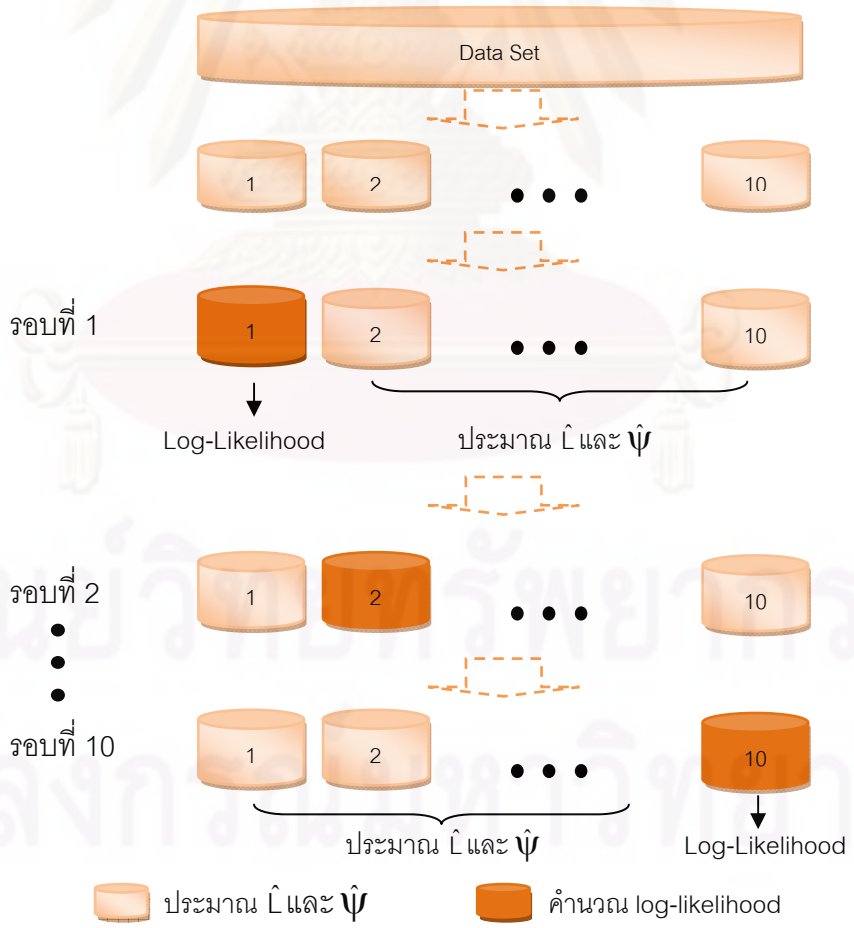
4. (10-fold) Likelihood Cross-Validation (LCV)

ครอสวาลิเดชั่น คือ วิธีการในการคาดการค่าความผิดพลาดของโมเดล หรือวิธีการที่เรานำเสนอ โดยพื้นฐานของวิธีครอสวาลิเดชั่น คือ การสุ่มตัวอย่าง (Resampling) โดย

เริ่มจากแบ่งชุดข้อมูลออกเป็นส่วนๆ และนำบางส่วนจากชุดข้อมูลนั้นมาตรวจสอบ ผลลัพธ์จากการทำครอสวาลิเดชั่นมักถูกใช้เป็นตัวเลือกในการกำหนดโมเดล

สำหรับการวิจัยครั้งนี้ใช้ครอสวาลิเดชั่นประเภท k-fold (k-fold Cross-Validation) คือ การแบ่งข้อมูลออกเป็น k ชุดเท่าๆ กัน และทำการคำนวณค่าความผิดพลาด k รอบ โดยแต่ละรอบการคำนวณ ข้อมูลชุดหนึ่งจากข้อมูล k ชุด จะถูกเลือกออกมาเพื่อเป็นข้อมูลทดสอบ และข้อมูลอีก k-1 ชุด จะถูกใช้เป็นข้อมูลสำหรับการเรียนรู้

ในที่นี้จะกำหนดให้ k=10 และวัดค่า log-likelihood สำหรับข้อมูลทดสอบ ดังนั้นจะเรียกเกณฑ์การคัดเลือกจำนวนปัจจัยสำหรับเกณฑ์นี้ว่า 10-fold Likelihood Cross-Validation คือ การแบ่งข้อมูลออกเป็น 10 ชุดเท่าๆ กัน และทำการคำนวณค่า log-likelihood 10 รอบ โดยแต่ละรอบการคำนวณ ข้อมูลชุดหนึ่งจากข้อมูล 10 ชุด จะถูกเลือกออกมาเพื่อเป็นข้อมูลสำหรับคำนวณค่า log-likelihood และข้อมูลอีก 9 ชุด จะถูกใช้เป็นข้อมูลสำหรับการประมาณค่า Factor Loading ($\hat{L}$) และ Specific Variance $(\hat{\psi})$ โดยมีขั้นตอนการทำงานของ 10-fold Likelihood Cross-Validation ดังแสดงในรูปที่ 2.3



รูปที่ 2.3 แสดงขั้นตอนการทำงานของ 10-fold Likelihood Cross-Validation

ให้ $\underset{\sim}{x}_i = (x_{1i}, x_{2i}, \ldots, x_{pi})^t$ เป็นเวกเตอร์ของค่าสังเกตจำนวน p ตัวแปร สำหรับค่าสังเกตค่าที่ i จากค่าสังเกตทั้งหมด n ค่า โดยมีข้อสมมติว่าเวกเตอร์สุ่ม $\underset{\sim}{X}$ เป็นเวกเตอร์สุ่มจากประชากรที่มีการแจกแจงแบบปกติหลายตัวแปร ด้วยเวกเตอร์ค่าเฉลี่ย $\underset{\sim}{\mu}$ และเมทริกซ์ความแปรปรวนร่วม Σ โดยให้ $S = \{1,2,...,n\}$ เป็นเซตของดัชนีค่าสังเกตทั้งหมด และกำหนดให้ตัวประมาณของ $\underset{\sim}{\mu}$ และ Σ ในการวิเคราะห์ปัจจัยที่มี m ปัจจัยร่วม ($1 \le m \le p$) คือ $\hat{\underset{\sim}{\mu}}(S)$ และ $\hat{\Sigma}(S;m)$ [8]

ให้ S_h คือ เซตของดัชนีของค่าสังเกตกลุ่มที่ h ซึ่งค่าสังเกตได้ถูกแบ่งออกเป็น k กลุ่ม (ในที่นี้จะพิจารณาเฉพาะ k=10) โดยที่ $1 \le h \le k$ และ k เป็นจำนวนกลุ่ม (k-fold) ดังนั้น $S_h^c = S \setminus S_h$ คือ เซตของดัชนีค่าสังเกตที่ไม่รวมดัชนีค่าสังเกตในกลุ่มที่ h และ $S_{h(i)}$ คือ เซตของดัชนีของค่าสังเกตกลุ่มที่ h ตัวที่ i ดังนั้นจะได้ค่า LCV สำหรับค่าสังเกตค่าที่ i คือ

$$LCV_i = L(\underset{\sim}{x}_i; \hat{\underset{\sim}{\mu}}(S^c_{h(i)}), \hat{\Sigma}(S^c_{h(i)}, m)) \tag{14}$$

เมื่อพิจารณา log-likelihood ในสมการ (14) จะได้ว่า

$$\ln(LCV_i) = -\frac{1}{2}(\underset{\sim}{x}_i - \hat{\underset{\sim}{\mu}}(S^c_{h(i)}))^t \cdot \hat{\Sigma}^{-1}(S^c_{h(i)};m) \cdot (\underset{\sim}{x}_i - \hat{\underset{\sim}{\mu}}(S^c_{h(i)})) - \frac{1}{2}\ln((2\pi)^p \cdot \left|\hat{\Sigma}(S^c_{h(i)};m)\right|) \tag{15}$$

สำหรับค่าสังเกต n_h ค่าของ S_h สมการ (15) จะกลายเป็น

$$\ln(LCV(S_h)) = -\frac{1}{2}\sum_{i=1}^{n_h}(\underset{\sim}{x}_i - \hat{\underset{\sim}{\mu}}(S^c_h))^t \cdot \hat{\Sigma}^{-1}(S^c_h;m) \cdot (\underset{\sim}{x}_i - \hat{\underset{\sim}{\mu}}(S^c_h)) - \frac{1}{2}n_h\ln((2\pi)^p \cdot \left|\hat{\Sigma}(S^c_h;m)\right|) \tag{16}$$

ดังนั้น log-likelihood สำหรับตัวแบบที่มี m ปัจจัย คือ

$$\ln(LCV(S;m)) = \sum_{h=1}^{10}\ln(LCV(S_h;m)) \tag{17}$$

<u>เกณฑ์การตัดสินใจ</u> คือ เลือกจำนวนปัจจัยที่มีค่า ln(LCV(S;m)) มากที่สุด

5. เกณฑ์ข้อสนเทศ (Information Criteria : IC)

เกณฑ์ข้อสนเทศเป็นการวัดระยะห่างระหว่าง 2 การแจกแจง โดยเกณฑ์ข้อสนเทศมีพื้นฐานอยู่บนข้อสนเทศของคูลล์แบค-ไลเบลอร์ (Kullback-Liebler Information: K-L Information) ซึ่งคูลล์แบค-ไลเบลอร์ได้เสนอข้อสนเทศดังกล่าวไว้ในปี ค.ศ. 1951 เพื่อเป็นการวัดความกลมกลืนระหว่างตัวแบบที่แท้จริงกับตัวแบบที่นำมาพิจารณา โดยกำหนดให้ [9]

$\mathbf{x}_n$ คือ เซตของค่าสังเกตทั้งหมด n ค่าที่ถูกสุ่มมาอย่างเป็นอิสระซึ่งกันและกันจากฟังก์ชันการแจกแจงความน่าจะเป็น (Probability Distribution Function) G(x) ที่ไม่ทราบการแจกแจงความน่าจะเป็น โดยที่ $\mathbf{x}_n = \{x_1, x_2, \ldots, x_n\}$

G(x) คือ ตัวแบบที่แท้จริงหรือการแจกแจงที่แท้จริง (True Model or True Distribution)

g(x) คือ ฟังก์ชันความหนาแน่น (Density Function) ของ G(x)

F(x) คือ ตัวแบบที่นำมาพิจารณาที่สร้างมาจากข้อมูล

f(x) คือ ฟังก์ชันความหนาแน่น (Density Function) ของ F(x)

$f(x|\hat{\theta})$ คือ ตัวแบบทางสถิติ (Statistical Model)

โดยข้อสนเทศของคูลล์แบค-ไลเบลอร์ คือ

$$I(G;F) = E_G\left[\ln\left\{\frac{G(X)}{F(X)}\right\}\right] \tag{18}$$

เมื่อ E_G คือ ค่าคาดหวังของการแจกแจงความน่าจะเป็น G

ถ้าฟังก์ชันการแจกแจงความน่าจะเป็นของตัวแปรสุ่มต่อเนื่องมีฟังก์ชันความหนาแน่น g(x) และ f(x) แล้วข้อสนเทศของคูลล์แบค-ไลเบลอร์ คือ

$$I(g;f) = \int_{-\infty}^{\infty} \ln\left\{\frac{g(x)}{f(x)}\right\} g(x)dx \tag{19}$$

และถ้าฟังก์ชันการแจกแจงความน่าจะเป็นของตัวแปรสุ่มไม่ต่อเนื่องมีฟังก์ชันความน่าจะเป็น $\{g(x_i); i = 1, 2, \ldots\}$ และ $\{f(x_i); i = 1, 2, \ldots\}$ แล้วข้อสนเทศของคูลล์แบค-ไลเบลอร์ คือ

$$I(g;f) = \sum_{i=1}^{\infty} g(x_i) \ln\left\{\frac{g(x_i)}{f(x_i)}\right\} \tag{20}$$

ดังนั้น ข้อสนเทศของคูลล์แบค-ไลเบลอร์สำหรับตัวแปรสุ่มต่อเนื่องและไม่ต่อเนื่อง คือ

$$I(g;f)=\int \ln\left\{\frac{g(x)}{f(x)}\right\}dG(x)$$

$$=\begin{cases}\int_{-\infty}^{\infty}\ln\left\{\dfrac{g(x)}{f(x)}\right\}g(x)dx, & \text{สำหรับการแจกแจงความน่าจะเป็นแบบต่อเนื่อง}\\ \sum_{i=1}^{\infty}g(x_i)\ln\left\{\dfrac{g(x_i)}{f(x_i)}\right\}, & \text{สำหรับการแจกแจงความน่าจะเป็นไม่แบบต่อเนื่อง}\end{cases} \quad (21)$$

คุณสมบัติของข้อสนเทศของคูลล์แบค-ไลเบลอร์ มีดังนี้

1. $I(g;f)\geq 0$
2. $I(g;f)=0 \Leftrightarrow g(x)=f(x)$

จากคุณสมบัติดังกล่าวของข้อสนเทศของคูลล์แบค-ไลเบลอร์ เราจึงพิจารณาค่าต่ำสุดของข้อสนเทศของคูลล์แบค-ไลเบลอร์ นั่นคือ ตัวแบบที่นำมาพิจารณา f(x) มีความใกล้เคียงตัวแบบที่แท้จริง g(x) แต่ในความเป็นจริงนั้น การคำนวณข้อสนเทศของคูลล์แบค-ไลเบลอร์มีข้อจำกัด เนื่องจากข้อสนเทศมีส่วนที่ต้องพิจารณาการแจกแจง g ซึ่งเราไม่ทราบการแจกแจงที่แท้จริง ดังนั้นจึงไม่สามารถคำนวณข้อสนเทศของคูลล์แบค-ไลเบลอร์ได้โดยตรง ดังรายละเอียดต่อไปนี้

$$I(G;F)=E_G\left[\ln\left\{\frac{G(X)}{F(X)}\right\}\right]$$

$$=E_G[\ln g(X)]-E_G[\ln f(X)] \quad (22)$$

เมื่อพิจารณาจากส่วนแรกของสมการ (22) จะพบว่า ขึ้นอยู่กับตัวแบบที่แท้จริง g เท่านั้น ดังนั้นจึงพิจารณาเฉพาะส่วนที่สองของสมการ (22) ก็เพียงพอต่อการที่จะเปรียบเทียบความแตกต่างระหว่าง 2 ตัวแบบ โดยจะเรียกส่วนที่สองนี้ว่า ค่าคาดหวังของ log-likelihood (Expected Log-Likelihood) สำหรับตัวแบบที่มีค่าคาดหวังของ log-likelihood มากจะเป็นตัวแบบที่ดี เนื่องจากทำให้ข้อสนเทศของคูลล์แบค-ไลเบลอร์มีค่าน้อย โดยค่าคาดหวังของ log-likelihood สำหรับตัวแบบที่มีการแจกแจงแบบต่อเนื่องและไม่ต่อเนื่อง คือ

$$E_G[\ln f(X)]=\int \ln f(x)dG(x)$$

$$=\begin{cases}\int_{-\infty}^{\infty}g(x)\ln f(x)dx, & \text{สำหรับการแจกแจงความน่าจะเป็นแบบต่อเนื่อง}\\ \sum_{i=1}^{\infty}g(x_i)\ln f(x_i), & \text{สำหรับการแจกแจงความน่าจะเป็นไม่แบบต่อเนื่อง}\end{cases} \quad (23)$$

จากสมการ (23) พบว่ายังขึ้นอยู่กับการแจกแจงที่แท้จริง g ซึ่งไม่ทราบการแจกแจงที่แท้จริงและยากที่จะอธิบายได้ชัดเจน อย่างไรก็ตามถ้าสามารถประมาณค่าคาดหวังของ log-likelihood ได้ดีจากข้อมูลค่าสังเกต ก็สามารถใช้ค่าประมาณดังกล่าวเป็นเกณฑ์ในการคัดเลือกตัวแบบได้ ซึ่งการประมาณค่าคาดหวังของ log-likelihood สามารถประมาณได้โดยการแทนที่การแจกแจงความน่าจะเป็นที่ไม่ทราบการแจกแจง G ด้วยฟังก์ชันการแจกแจงที่ได้จากการทดลอง (Empirical Distribution Function) $\hat{G}$ ซึ่งอยู่บนพื้นฐานของข้อมูล โดยฟังก์ชันการแจกแจงที่ได้จากการทดลองสำหรับฟังก์ชันความน่าจะเป็น $\hat{g}(x_i)=\frac{1}{n}$ โดยที่ i=1, 2, …, n นั่นคือ แต่ละค่าสังเกตมีความน่าจะเป็นเท่ากัน คือ $\frac{1}{n}$ ดังนั้น

$$E_{\hat{G}}=[\ln f(X)]=\int \ln f(x)d\hat{G}(x)$$

$$=\sum_{i=1}^{n}\hat{g}(x_i)\ln f(x_i)$$

$$=\frac{1}{n}\sum_{i=1}^{n}\ln f(x_i) \qquad (24)$$

จากกฎของเลขจำนวนมาก (Law of Large Numbers) เมื่อจำนวนของค่าสังเกต (n) มีมากพอสมควร แล้วค่าเฉลี่ยของตัวแปรสุ่ม $Y_i=\ln f(X_i);\ i=1,2,\ldots,n$ จะลู่เข้าในความน่าจะเป็นสู่ค่าคาดหวัง นั่นคือ

$$\frac{1}{n}\sum_{i=1}^{n}\ln f(X_i)\rightarrow E_G[\ln f(X)], \qquad n\rightarrow\infty$$

ดังนั้น เป็นที่ชัดเจนว่าการประมาณที่ขึ้นกับฟังก์ชันการแจกแจงที่ได้จากการทดลองในสมการ (24) เป็นการประมาณโดยธรรมชาติของค่าคาดหวังของ log-likelihood ดังนั้นค่าประมาณของค่าคาดหวังของ log-likelihood คูณด้วย n คือ

$$n\int \ln f(x)d\hat{G}(x)=\sum_{i=1}^{n}\ln f(x_i) \qquad (25)$$

สมการ (25) คือ log-likelihood ของตัวแบบ f(x)

การคัดเลือกตัวแบบโดยใช้เกณฑ์ข้อสนเทศเป็นการพิจารณาความกลมกลืนของตัวแบบทางสถิติที่นำมาพิจารณากับตัวแบบที่แท้จริง โดยยึดหลักของการพยากรณ์ นั่นคือการคำนวณค่าคาดหวังของความกลมกลืนของตัวแบบที่พิจารณา $f(z|\hat{\theta})$ ซึ่งเป็นตัวแบบที่ใช้ใน

การพยากรณ์ข้อมูลในอนาคต Z=z อย่างเป็นอิสระซึ่งกันและกัน โดยที่ Z=z ถูกสร้างมากจากการแจกแจงที่ แท้จริง g(z) ดังนั้นข้อสนเทศของคูลแบค-ไลเบลอร์ คือ

$$I\{g(z); f(z|\hat{\theta})\} = E_G\left[\ln\left\{\frac{g(Z)}{f(Z|\hat{\theta})}\right\}\right]$$

$$= E_G[\ln g(Z)] - E_G[\ln f(Z|\hat{\theta})] \quad (26)$$

จากคุณสมบัติของข้อสนเทศของคูลแบค-ไลเบลอร์ จะพิจารณาเฉพาะ $E_G[\ln f(Z|\hat{\theta})]$ นั่นคือ

$$E_G[\ln f(Z|\hat{\theta})] = \int \ln f(z|\hat{\theta}) d\hat{G}(z) \quad (27)$$

ซึ่งตัวแบบที่มีค่า $E_G[\ln f(Z|\hat{\theta})]$ มาก จะมีข้อสนเทศของคูลแบค-ไลเบลอร์น้อย จึงเป็นตัวแบบที่ใกล้เคียงกับตัวแบบที่แท้จริง ดังนั้นในการนิยามเกณฑ์ข้อสนเทศจะมีหลักที่สำคัญ คือ การทำให้ได้ตัวประมาณที่ดีของ $E_G[\ln f(Z|\hat{\theta})]$ ซึ่งตัวประมาณของ $E_G[\ln f(Z|\hat{\theta})]$ คือ

$$E_{\hat{G}}[\ln f(Z|\hat{\theta})] = \int \ln f(z|\hat{\theta}) d\hat{G}(z)$$

$$= -\frac{1}{n}\sum_{i=1}^{n} \ln f(x_i|\hat{\theta}) \quad (28)$$

ซึ่งตัวประมาณของ $E_G[\ln f(Z|\hat{\theta})]$ คือ การแทนการแจกแจงความน่าจะเป็น G ที่ไม่ทราบการแจกแจงด้วยฟังก์ชันการแจกแจงที่ได้จากการทดลอง $\hat{G}$ และ log-likelihood ของตัวแบบ $f(z|\hat{\theta})$ คือ

$$\ell(\hat{\theta}) = \sum_{i=1}^{n} \ln f(x_i|\hat{\theta}) \quad (29)$$

ดังนั้นจึงสรุปได้ว่า ตัวประมาณของ $E_G[\ln f(Z|\hat{\theta})]$ คือ $\frac{1}{n}\ell(\hat{\theta})$ และ $\ell(\hat{\theta})$ คือ ตัวประมาณของ $nE_G[\ln f(Z|\hat{\theta})]$

จากการที่เรามีข้อมูลค่าสังเกตอย่างจำกัด ดังนั้นจึงได้สร้างตัวแบบทางสถิติหลายๆ ตัวแบบจากข้อมูลที่มีอยู่ เพื่อนำมาพิจารณาเลือกตัวแบบที่ใกล้เคียงตัวแบบที่แท้จริงมากที่สุด โดยตัวแบบที่นำมาพิจารณาคือ $\{f_j(z|\theta_j); j=1,2,\ldots,r\}$ โดยที่ $\hat{\theta}_j$ คือ ตัวประมาณของพารามิเตอร์ θ_j ซึ่งการคัดเลือกตัวแบบที่เหมาะสมจะพิจารณาจากขนาดของ $\ell(\hat{\theta})$ แต่ $\ell(\hat{\theta})$ มีความเอนเอียงอยู่ เนื่องจากข้อมูลถูกนำมาใช้ในการประมาณพารามิเตอร์ของตัวแบบและการ

ประมาณค่าคาดหวังของ log-likelihood ($E_G[\ln f(Z|\hat{\theta})]$) ดังนั้น จึงต้องคำนวณค่าความเอนเอียงเพื่อถ่วงน้ำหนักของ $\ell(\hat{\theta})$ โดยกำหนดให้

$\mathbf{x_n} = \{x_1, x_2, \ldots, x_n\}$ เป็นค่าสังเกต n ค่า ที่ถูกสร้างมาจากการแจกแจงที่แท้จริง G(x) หรือ g(x) ที่มีตัวแปรสุ่ม $\mathbf{X_n} = (X_1, X_2, \ldots, X_n)^t$ และให้

$$\ell(\hat{\theta}) = \sum_{i=1}^{n} \ln f(x_i | \hat{\theta}(\mathbf{x_n})) = \ln f(\mathbf{x_n} | \hat{\theta}(\mathbf{x_n})) \quad (30)$$

เป็น log-likelihood ของตัวแบบทางสถิติ $f(z|\hat{\theta}(\mathbf{x_n}))$ โดยที่ความเอนเอียงของ log-likelihood คือ

$$b(G) = E_{G(\mathbf{x_n})}[\ln f(\mathbf{X_n} | \hat{\theta}(\mathbf{X_n})) - nE_{G(z)}[\ln f(Z | \hat{\theta}(\mathbf{X_n})]] \quad (31)$$

เมื่อ $E_{G(\mathbf{x_n})}$ คือ ค่าคาดหวังของการแจกแจงร่วม (Joint Distribution) $\prod_{i=1}^{n} G(x_i) = G(\mathbf{x_n})$ ของ $\mathbf{X_n}$

$E_{G(z)}$ คือ ค่าคาดหวังของการแจกแจงที่แท้จริง G(z)

ดังนั้นรูปทั่วไปของเกณฑ์สนเทศ (Information Criteria) คือ

$IC(\mathbf{X_n}; \hat{G})$ = -2(log-likelihood ของตัวแบบทางสถิติ - ตัวประมาณของความเอนเอียง)

$$= -2\sum_{i=1}^{n} \ln f(X_i | \hat{\theta}) + 2 \text{ (ตัวประมาณของ } b(G)) \quad (32)$$

เกณฑ์การคัดเลือกจำนวนปัจจัยในการวิเคราะห์ปัจจัยที่ใช้ในการวิจัยครั้งนี้ ได้แก่

5.1 เกณฑ์การคัดเลือกจำนวนปัจจัยโดยใช้ข้อสนเทศของอากาอิเกะ (Akaike's Information Criteria: AIC)

ในปี ค.ศ.1973 อากาอิเกะได้นำเสนอเกณฑ์ที่ใช้ในการคัดเลือกตัวแบบเพื่อใช้เป็นเครื่องมือในการหาตัวแบบที่ให้ค่าพยากรณ์ที่แม่นยำที่สุด และได้มาจากค่าคาดหวังของ log-likelihood ซึ่งเป็นเกณฑ์พื้นฐาน นั่นคือ เกณฑ์การคัดเลือกตัวแบบโดยใช้ข้อสนเทศของอากาอิเกะ (AIC) ซึ่งเป็นเกณฑ์การคัดเลือกตัวแบบที่พิจารณาจากการประมาณความคลาดเคลื่อนรวมกับข้อสนเทศของค่าสังเกต โดยใช้แนวคิดจากค่าต่ำสุดของข้อสนเทศของคูลแบค-ไลเบลอร์ (Kullback-Leibler Information : K-L) ซึ่งข้อสนเทศของคูลแบค-ไลเบลอร์ ใช้เป็นเครื่องมือในการวัดระยะทางระหว่างตัวแบบที่แท้จริง (True Model) กับตัวแบบที่นำมาพิจารณา (Candidate

Model) ดังนั้นระยะทางดังกล่าวควรมีค่าต่ำที่สุด กล่าวคือ AIC เป็นตัวประมาณที่ไม่เอนเอียงโดยประมาณของค่าคาดหวังของข้อสนเทศของคูลแบค-ไลเบลอร์ โดยที่ [9]

$$AIC = -2\ell(\hat{\theta}) + 2K \tag{33}$$

เมื่อ K คือ พารามิเตอร์ที่เป็นอิสระ (Free Parameter)

สำหรับการใช้งาน AIC ในการวิเคราะห์ปัจจัยนั้น

$$\ell(\hat{\theta};m) = -\frac{1}{2}n\ln\left((2\pi)^p \cdot \left|\hat{\Sigma}(m)\right|\right) - \frac{1}{2}\sum_{i=1}^{n}(\underset{\sim}{x}_j - \hat{\underset{\sim}{\mu}})^t \cdot \hat{\Sigma}^{-1}(m) \cdot (\underset{\sim}{x}_j - \hat{\underset{\sim}{\mu}})$$

$$= -\frac{1}{2}n\ln\left((2\pi)^p \left|\hat{L}\hat{L}^t + \hat{\Psi}\right|\right) - \frac{1}{2}\sum_{i=1}^{n}(\underset{\sim}{x}_j - \hat{\underset{\sim}{\mu}})^t \cdot (\hat{L}\hat{L}^t + \hat{\Psi})^{-1} \cdot (\underset{\sim}{x}_j - \hat{\underset{\sim}{\mu}}) \tag{34}$$

เมื่อ $\hat{\Sigma}(m) = \hat{L}\hat{L}^t + \hat{\Psi}$ คือ เมทริกซ์ความแปรปรวนร่วมสำหรับ m ปัจจัย ในการวิเคราะห์ปัจจัย

และ $K = p(m+1) - \frac{1}{2}m(m-1)$ ดังนั้น

$$AIC(m) = -2\ell(\hat{\theta};m) + 2\left(p(m+1) - \frac{1}{2}m(m-1)\right) \tag{35}$$

<u>เกณฑ์การตัดสินใจ</u> คือ เลือกจำนวนปัจจัยที่มีค่า AIC ต่ำที่สุด

5.2 เกณฑ์การคัดเลือกจำนวนปัจจัยโดยใช้ข้อสนเทศของชวาร์ช (Schwarz's Information Criteria : SIC)

หรือเกณฑ์ข้อสนเทศของเบส์ (Bayesian Information Criteria: BIC) ชวาร์ชได้เสนอไว้ในปี ค.ศ.1978 โดยนำแนวคิดของเบส์มาประยุกต์ใช้ โดยสมมติว่าความน่าจะเป็นโดยหลักเกณฑ์ (Prior Probability) ของทุกตัวแบบเหมือนกัน ดังนั้นในการพิจารณาการคัดเลือกตัวแบบที่เหมาะสมจึงพิจารณาจากความน่าจะเป็นโดยประสบการณ์ (Posterior Probability) คือ พิจารณาตัวแบบที่มีความน่าจะเป็นโดยประสบการณ์สูงที่สุดหรือพิจารณาตัวแบบที่ให้ค่า SIC ต่ำที่สุดเป็นตัวแบบที่ดีที่สุด [9]

เกณฑ์การคัดเลือกตัวแบบโดยใช้ข้อสนเทศของชวาร์ชได้นำแนวคิดของเบส์มาประยุกต์ดังรายละเอียดต่อไปนี้

กำหนดให้

$\mathbf{x}_n$ คือ เซตของข้อมูลค่าสังเกต โดยที่ $\mathbf{x}_n = \{x_1, x_2, \ldots, x_n\}$

M_i คือ ตัวแบบที่นำมาพิจารณา (Candidate Model) โดยที่ i=1, 2, …, r

$f_i(x|\theta_i)$ คือ การแจกแจงของ x เมื่อกำหนดเวกเตอร์พารามิเตอร์ θ_i โดยที่ $\theta_i \in \Theta_i \subset R^{k_i}$

$\pi_i(\theta_i)$ คือ ความน่าจะเป็นโดยหลักเกณฑ์ (Prior Probability) ของเวกเตอร์พารามิเตอร์ θ_i ที่มี k_i มิติ

$p_i(\mathbf{x}_n)$ คือ ความน่าจะเป็นส่วนริม (Marginal Probability) หรือความน่าจะเป็นของ $\mathbf{x}_n$ โดยที่

$$p_i(\mathbf{x}_n) = \int f_i(x_n \mid \theta_i)\pi_i(\theta_i)d\theta_i \qquad (36)$$

ซึ่งความน่าจะเป็นส่วนริมนี้สามารถพิจารณาได้จากภาวะน่าจะเป็น (Likelihood) ของตัวแบบที่ i (M_i) และเรียกภาวะน่าจะเป็นนี้ว่า ภาวะน่าจะเป็นส่วนริม (Marginal Likelihood) และจากทฤษฎีบทของเบส์ (Bayes' Theorm) กำหนดให้

$P(M_i)$ คือ ความน่าจะเป็นโดยหลักเกณฑ์ของตัวแบบที่ i

$P(M_i|\mathbf{x}_n)$ คือ ความน่าจะเป็นโดยประสบการณ์ของตัวแบบที่ i โดยที่

$$P(M_i \mid \mathbf{x}_n) = \frac{p_i(\mathbf{x}_n)P(M_i)}{\sum_{j=1}^{r} p_j(\mathbf{x}_n)P(M_j)}, \qquad i = 1, 2, \ldots, r \qquad (37)$$

ซึ่งความน่าจะเป็นโดยประสบการณ์นี้เป็นตัววัดความน่าจะเป็นของข้อมูลที่ถูกสร้างจากตัวแบบที่ i เมื่อมี $\mathbf{x}_n$ ดังนั้น ถ้ามีตัวแบบหนึ่งถูกเลือกมาจาก r ตัวแบบที่นำมาพิจารณา นั่นหมายความว่า ตัวแบบนั้นมีความน่าจะเป็นโดยประสบการณ์มากที่สุด ด้วยหลักเกณฑ์นี้ตัวแบบที่มีค่า $p_i(\mathbf{x}_n)P(M_i)$ มากที่สุดจึงเป็นตัวแบบที่ถูกเลือก เนื่องจากทุกตัวแบบมี $\sum_{i=1}^{r} p_j(\mathbf{x}_n)P(M_j)$ เหมือนกัน

ถ้าเราเพิ่มข้อสมมติว่า ความน่าจะเป็นโดยหลักเกณฑ์ $P(M_i)$ เท่ากันในทุกตัวแบบ สิ่งที่ได้ตามมา คือ ตัวแบบที่มีภาวะน่าจะเป็นส่วนริม $p_i(\mathbf{x}_n)$ มากที่สุดจะเป็นตัวแบบที่ถูกเลือก ดังนั้น จึงต้องทำการประมาณภาวะน่าจะเป็นส่วนริมในสมการ (36) โดยภาวะน่าจะเป็นส่วนริมหรือความน่าจะเป็นส่วนริม (Marginal Distribution) ของ $\mathbf{x}_n$ สามารถประมาณได้ด้วยวิธีลาปลาซ (Laplace's Method) ดังนี้ (ในส่วนต่อไปนี้จะละสัญลักษณ์ของตัวแบบที่ i (M_i) ไว้)

$$\begin{aligned} p(\mathbf{x}_n) &= \int f(\mathbf{x}_n \mid \theta)\pi(\theta)d\theta \\ &= \int \exp\{\ln f(\mathbf{x}_n \mid \theta)\}\pi(\theta)d\theta \\ &= \int \exp\{\ell(\theta)\}\pi(\theta)d\theta \end{aligned} \qquad (38)$$

เมื่อ θ คือ เวกเตอร์ของพารามิเตอร์ที่มี K มิติ

$\ell(\theta)$ คือ ฟังก์ชันภาวะน่าจะเป็น $\ell(\theta) = \ln f(\mathbf{x}_n \mid \theta)$

จากอนุกรมเทย์เลอร์ของฟังก์ชันภาวะน่าจะเป็น $\ell(\theta)$ รอบ $\hat{\theta}$ คือ

$$\ell(\theta)=\ell(\hat{\theta})-\frac{n}{2}(\theta-\hat{\theta})^{t}J(\hat{\theta})(\theta-\hat{\theta})+\ldots \tag{39}$$

เมื่อ $J(\hat{\theta})=-\frac{1}{n}\frac{\partial^{2}\ell(\theta)}{\partial\theta\partial\theta^{t}}\Big|_{\theta=\hat{\theta}}=-\frac{1}{n}\frac{\partial^{2}\ln f(x_{n}\mid\theta)}{\partial\theta\partial\theta^{t}}\Big|_{\theta=\hat{\theta}}$

ในทำนองเดียวกัน เราสามารถขยายการแจกแจงโดยหลักเกณฑ์ $\pi(\theta)$ ด้วยอนุกรมเทย์เลอร์รอบ $\hat{\theta}$ ได้ดังนี้

$$\pi(\theta)=\pi(\hat{\theta})+(\theta-\hat{\theta})^{t}\frac{\partial\pi(\theta)}{\partial\theta}\Big|_{\theta=\hat{\theta}}+\ldots \tag{40}$$

ดังนั้นเมื่อแทนสมการ (39) และสมการ (40) ในสมการ (38) แล้วทำให้ได้ค่าประมาณของภาวะน่าจะเป็นส่วนริม ดังนี้

$$p(x_{n})=\int\exp\left\{\ell(\hat{\theta})-\frac{n}{2}(\theta-\hat{\theta})^{t}J(\hat{\theta})(\theta-\hat{\theta})+\ldots\right\}\times\left\{\pi(\hat{\theta})+(\theta-\hat{\theta})^{t}\frac{\partial\pi(\theta)}{\partial\theta}\Big|_{\theta=\hat{\theta}}\right\}d\theta$$

$$\approx\exp\{\ell(\hat{\theta})\}\pi(\hat{\theta})\int\exp\left\{-\frac{n}{2}(\theta-\hat{\theta})^{t}J(\hat{\theta})(\theta-\hat{\theta})\right\}d\theta \tag{41}$$

จากการที่ $\hat{\theta}$ ลู่เข้าสู่ θ ในความน่าจะเป็น เมื่อ $\hat{\theta}-\theta=O_{p}(n^{-1/2})$ ดังนั้น

$$\int(\theta-\hat{\theta})\exp\left\{-\frac{n}{2}(\theta-\hat{\theta})^{t}J(\hat{\theta})(\theta-\hat{\theta})\right\}d\theta=0 \tag{42}$$

เพราะฉะนั้นจะได้

$$\int\exp\left\{-\frac{n}{2}(\theta-\hat{\theta})^{t}J(\theta)(\theta-\hat{\theta})\right\}d\theta=(2\pi)^{K/2}n^{-K/2}\left|J(\hat{\theta})\right|^{-1/2} \tag{43}$$

เนื่องจากอินทิแกรนด์ (Integrand) ข้างต้นนี้ คือ ฟังก์ชันความหนาแน่นของการแจกแจงแบบปกติ (Normal Distribution) K มิติ ที่มีเวกเตอร์ค่าเฉลี่ย $\hat{\theta}$ และเมทริกซ์ความแปรปรวนร่วม $\frac{J^{-1}(\hat{\theta})}{n}$ ดังนั้นเมื่อขนาดตัวอย่างเพิ่มขึ้น ภาวะน่าจะเป็นส่วนริมจะสามารถประมาณได้ด้วย

$$p(x_{n})\approx\exp\{\ell(\hat{\theta})\}\pi(\hat{\theta})(2\pi)^{K/2}n^{-K/2}\left|J(\hat{\theta})\right|^{-1/2} \tag{44}$$

เพราะฉะนั้นเมื่อพิจารณาลอการิทึมของสมการ (44) คูณด้วย -2 จะได้

$$-2\ln p(\mathbf{x}_n) = -2\ln\{\int f(x_n \mid \theta)\pi(\theta)d\theta\}$$

$$\approx -2\ell(\hat{\theta}) + K\ln n + \ln\left|J(\hat{\theta})\right| - p\ln(2\pi) - 2\ln\pi(\hat{\theta}) \qquad (45)$$

สามารถละสามเทอมหลังได้ เนื่องจากมีค่าน้อยกว่า $O(1)$ เมื่อพิจารณาในส่วนของขนาดตัวอย่าง ดังนั้นเกณฑ์ข้อสนเทศของชวาร์ซจึงกลายเป็น

$$SIC = -2\ell(\hat{\theta}) + K\ln n \qquad (46)$$

เมื่อ K คือ พารามิเตอร์ที่เป็นอิสระ

สำหรับการใช้งาน SIC ในการวิเคราะห์ปัจจัยนั้น

$$\ell(\hat{\theta}; m) = -\frac{1}{2}n\ln\left((2\pi)^p \cdot \left|\hat{\Sigma}(m)\right|\right) - \frac{1}{2}\sum_{i=1}^{n}(\underset{\sim}{x}_i - \underset{\sim}{\hat{\mu}})^t \cdot \hat{\Sigma}^{-1}(m) \cdot (\underset{\sim}{x}_i - \underset{\sim}{\hat{\mu}})$$

$$= -\frac{1}{2}n\ln\left((2\pi)^p \left|\hat{L}\hat{L}^t + \hat{\Psi}\right|\right) - \frac{1}{2}\sum_{i=1}^{n}(\underset{\sim}{x}_i - \underset{\sim}{\hat{\mu}})^t \cdot (\hat{L}\hat{L}^t + \hat{\Psi})^{-1} \cdot (\underset{\sim}{x}_i - \underset{\sim}{\hat{\mu}}) \qquad (47)$$

เมื่อ $\hat{\Sigma}(m) = \hat{L}\hat{L}^t + \hat{\Psi}$ คือ เมทริกซ์ความแปรปรวนร่วมสำหรับ m ปัจจัย ในการวิเคราะห์ปัจจัย และ $K = p(m+1) - \frac{1}{2}m(m-1)$ และ n คือ ขนาดตัวอย่าง ดังนั้น

$$SIC(m) = -2\ell(\hat{\theta}; m) + \left(p(m+1) - \frac{1}{2}m(m-1)\right)\ln(n) \qquad (48)$$

<u>เกณฑ์การตัดสินใจ</u> คือ เลือกจำนวนปัจจัยที่มีค่า SIC ต่ำที่สุด

5.3 เกณฑ์การคัดเลือกจำนวนปัจจัยโดยใช้ข้อสนเทศของแฮนแนนและควินน์ (Hannan and Quinn's Information Criteria : HQ)

ในปี ค.ศ. 1979 แฮนแนนและควินน์ (Hannan and Quinn) ได้เสนอเกณฑ์การคัดเลือกตัวแบบสำหรับการวิเคราะห์อนุกรมเวลาไว้ โดยเป็นเกณฑ์ที่มีรูปแบบคล้ายกับเกณฑ์ข้อสนเทศของอากาอิเกะ แต่ยังคงคุณสมบัติความคงเส้นคงวา (Consistency) อยู่ โดยมีรูปทั่วไปดังนี้

$$HQ = -2\ell(\hat{\theta}) + 2KC\ln(\ln(n)) \qquad (49)$$

เมื่อ C > 1

ต่อมาแฮนแนนและควินน์พบว่า เกณฑ์ HQ ยังคงคุณสมบัติคงเส้นคงวาแม้ C=1 ดังนั้นเกณฑ์ข้อสนเทศของแฮนแนนและควินน์ที่นิยมใช้ทั่วไป จึงมีรูปแบบดังนี้

$$HQ = -2\ell(\hat{\theta}) + 2K\ln(\ln(n)) \tag{50}$$

สำหรับการใช้งาน HQ ในการวิเคราะห์ปัจจัยนั้น

$$\ell(\hat{\theta};m) = -\frac{1}{2}n\ln\left((2\pi)^p \cdot \left|\hat{\Sigma}(m)\right|\right) - \frac{1}{2}\sum_{i=1}^{n}(\underset{\sim}{x}_i - \hat{\underset{\sim}{\mu}})^t \cdot \hat{\Sigma}^{-1}(m) \cdot (\underset{\sim}{x}_i - \hat{\underset{\sim}{\mu}})$$

$$= -\frac{1}{2}n\ln\left((2\pi)^p \left|\hat{L}\hat{L}^t + \hat{\Psi}\right|\right) - \frac{1}{2}\sum_{i=1}^{n}(\underset{\sim}{x}_i - \hat{\underset{\sim}{\mu}})^t \cdot (\hat{L}\hat{L}^t + \hat{\Psi})^{-1} \cdot (\underset{\sim}{x}_i - \hat{\underset{\sim}{\mu}}) \tag{51}$$

เมื่อ $\hat{\Sigma}(m) = \hat{L}\hat{L}^t + \hat{\Psi}$ คือ เมทริกซ์ความแปรปรวนร่วมสำหรับ m ปัจจัยในการวิเคราะห์ปัจจัย และ $K = p(m+1) - \frac{1}{2}m(m-1)$ และ n คือ ขนาดตัวอย่าง ดังนั้น

$$HQ(m) = -2\ell(\hat{\theta};m) + 2\left(p(m+1) - \frac{1}{2}m(m-1)\right)\ln(\ln(n)) \tag{52}$$

<u>เกณฑ์การตัดสินใจ</u> คือ เลือกจำนวนปัจจัยที่มีค่า HQ ต่ำที่สุด

6. การทดสอบสมมติฐานผลต่างระหว่างค่าสัดส่วนของประชากร 2 กลุ่ม

เป็นการทดสอบเกี่ยวกับการเปรียบเทียบสัดส่วนของลักษณะที่สนใจศึกษาของประชากร 2 กลุ่ม ว่าแตกต่างกันหรือไม่ หรือสัดส่วนของลักษณะที่สนใจศึกษาของประชากรกลุ่มที่ 1 มากกว่าหรือน้อยกว่าประชากรกลุ่มที่ 2 โดยกำหนดให้ p_1 และ p_2 แทนสัดส่วนของลักษณะที่สนใจศึกษาของประชากรกลุ่มที่ 1 และประชากรกลุ่มที่ 2 ตามลำดับ ดังนั้น $\hat{p}_1$ และ $\hat{p}_2$ แทนสัดส่วนตัวอย่างของลักษณะที่สนใจศึกษาที่สุ่มจากประชากรกลุ่มที่ 1 และประชากรกลุ่มที่ 2 ตามลำดับ และเนื่องจากการประมาณค่าผลต่างระหว่างสัดส่วนประชากร 2 กลุ่ม จะใช้ตัวอย่างขนาดใหญ่ ดังนั้น $\hat{p}_1 - \hat{p}_2$ จึงมีการแจกแจงแบบปกติที่มีค่าเฉลี่ยเท่ากับ $p_1 - p_2$ และค่าแปรปรวนเท่ากับ $\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}$ เพราะฉะนั้นในการทดสอบสมมติฐานจึงใช้สถิติทดสอบ Z โดยที่สมมติฐานที่ใช้ในการทดสอบ คือ [5]

$$H_0 : p_1 - p_2 \le p_0$$
$$H_1 : p_1 - p_2 > p_0$$

โดยที่ p_0 แทน ค่าคงที่ผลต่างระหว่าง p_1 และ p_2 ภายใต้สมมติฐาน $(p_0 = p_1 - p_2)$

ตัวสถิติทดสอบ คือ $$Z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sigma_{\hat{p}_1 - \hat{p}_2}}$$

$$= \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\left(\frac{p_1(1-p_1)}{n_1}\right) + \left(\frac{p_2(1-p_2)}{n_2}\right)}}$$

เนื่องจากไม่ทราบค่า p_1 และ p_2 จึงประมาณ $\sigma_{\hat{p}_1 - \hat{p}_2}$ ด้วย $\hat{\sigma}_{\hat{p}_1 - \hat{p}_2}$ โดยที่

$$\hat{\sigma}_{\hat{p}_1 - \hat{p}_2} = \sqrt{\left(\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}\right) + \left(\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}\right)}$$

ดังนั้น

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\left(\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}\right) + \left(\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}\right)}}$$

โดยที่ขอบเขตปฏิเสธสมมติฐานหลัก คือ $Z > Z_{1-\alpha}$ หรือ $\frac{p-\text{value}}{2} < \alpha$

แต่ถ้า $p_0 = 0$ ทำให้ $\hat{\sigma}_{\hat{p}_1 - \hat{p}_2} = \sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$ โดยที่ $\hat{p} = \frac{\sum_{i=1}^{n_1} X_{1i} + \sum_{i=1}^{n_2} X_{2i}}{n_1 + n_2}$

ดังนั้น

$$Z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

บทที่ 3

วิธีดำเนินการวิจัย

การศึกษาวิจัยในครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพเกณฑ์การคัดเลือกจำนวนปัจจัยในการวิเคราะห์ปัจจัย โดยเกณฑ์ที่ใช้ในการเปรียบเทียบ คือ

1. (10-fold) Likelihood Cross-Validation (LCV)
2. เกณฑ์การคัดเลือกจำนวนปัจจัยโดยใช้ข้อสนเทศของอากาอิเกะ (AIC)
3. เกณฑ์การคัดเลือกจำนวนปัจจัยโดยใช้ข้อสนเทศของชวาร์ซ (SIC)
4. เกณฑ์การคัดเลือกจำนวนปัจจัยโดยใช้ข้อสนเทศของแฮนแนนและควินน์ (HQ)

โดยพิจารณาค่าอัตราความถูกต้อง (%) ของการคัดเลือกจำนวนปัจจัยที่ถูกต้องของเกณฑ์การคัดเลือกปัจจัยทั้ง 4 เกณฑ์ข้างต้น เป็นเกณฑ์การตัดสินใจ โดยใช้โปรแกรม R 2.9.0 ในการประมวลผล ซึ่งมีขั้นตอนของการดำเนินการวิจัยดังนี้

แผนการศึกษาวิจัย

ผู้วิจัยได้กำหนดสถานการณ์ต่างๆ สำหรับการวิจัยครั้งนี้ไว้ดังนี้

1. กำหนดให้จำนวนตัวแปรเท่ากับ 10 ตัวแปร จำนวนปัจจัยเท่ากับ 1,2,...,5 ปัจจัย และขนาดตัวอย่างเท่ากับ 200, 300, 500 และ 1,000
2. กำหนดให้จำนวนตัวแปรเท่ากับ 20 ตัวแปร จำนวนปัจจัยเท่ากับ 1,2,...,10 ปัจจัย และขนาดตัวอย่างเท่ากับ 300, 500 และ 1,000
3. กำหนดให้จำนวนตัวแปรเท่ากับ 30 ตัวแปร จำนวนปัจจัยเท่ากับ 1,2,...,15 ปัจจัย และขนาดตัวอย่างเท่ากับ 500 และ 1,000
4. กำหนดให้จำนวนตัวแปรเท่ากับ 40 ตัวแปร จำนวนปัจจัยเท่ากับ 1,2,...,20 ปัจจัย และขนาดตัวอย่างเท่ากับ 1,000

ขั้นตอนในการดำเนินการวิจัย

ขั้นตอนในการดำเนินการวิจัยมีดังนี้

1. กำหนดลักษณะของข้อมูลตามจำนวนตัวแปร (p) จำนวนปัจจัย (m) และขนาดตัวอย่าง (n) ที่กำหนดในขอบเขตการวิจัย

2. สร้างเมทริกซ์สหสัมพันธ์ Σ_p ด้วยวิธี Onion เพื่อสร้างเมทริกซ์ความแปรปรวนร่วม (Covariance Matrix : Σ) ที่มีค่าแปรปรวน (Variance) เท่ากับ 1 สำหรับตัวแปร X จำนวน p ตัวแปร ซึ่งจะทำให้ได้ Factor Loading สำหรับ m ปัจจัย และเมทริกซ์ความแปรปรวนของค่าเฉพาะ (ψ) (ทำให้สามารถสร้างค่าเฉพาะ (ε) ได้)

3. สร้างข้อมูลของตัวแปร X จำนวน p ตัวแปร ที่มีการแจกแจงแบบปกติหลายตัวแปร (Multivariate Normal Distribution) มีเวกเตอร์ค่าเฉลี่ยเท่ากับ $\underset{\sim}{0}$ และเมทริกซ์ความแปรปรวนร่วม Σ ที่ได้จากข้อ 2 ($X \sim N_p(\underset{\sim}{0}, \Sigma)$) จากตัวแบบปัจจัยตามสมการ (3) โดยมีจำนวนปัจจัย m ปัจจัยตามที่กำหนดในข้อ 1

4. สมมติว่าเราไม่ทราบค่า m จึงทำการคัดเลือกจำนวนปัจจัยที่เหมาะสมด้วยเกณฑ์ทั้ง 4 เกณฑ์ โดยกำหนดให้ $\hat{m}$ แทน จำนวนปัจจัยที่ประมาณได้ ซึ่งเกณฑ์ทั้ง 4 เกณฑ์ ได้แก่

4.1 10-fold Likelihood Cross-Validation (LCV)

4.2 เกณฑ์การคัดเลือกจำนวนปัจจัยโดยใช้ข้อสนเทศของอากาอิเกะ (AIC)

4.3 เกณฑ์การคัดเลือกจำนวนปัจจัยโดยใช้ข้อสนเทศของชวาร์ซ (SIC)

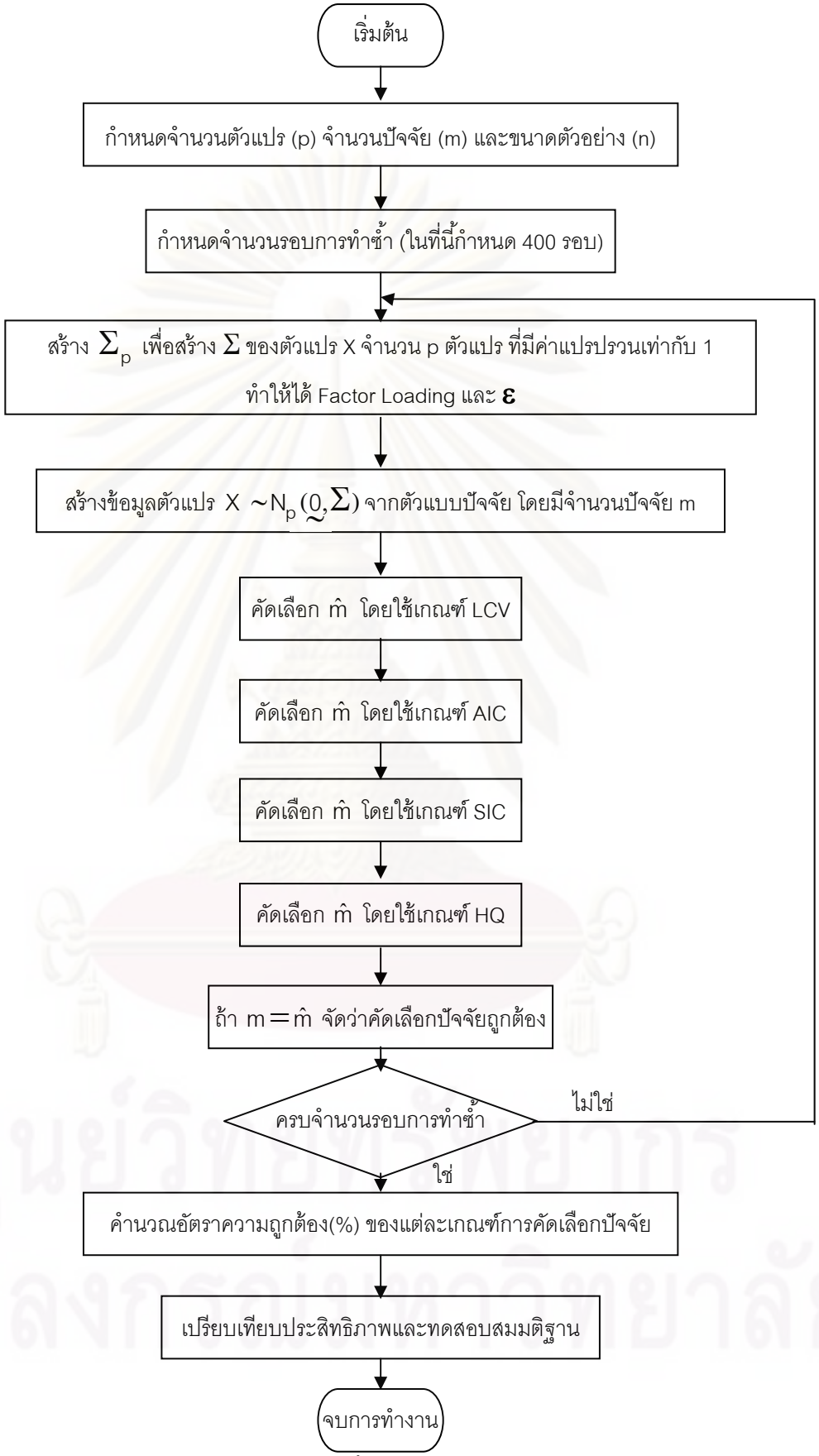
4.4 เกณฑ์การคัดเลือกจำนวนปัจจัยโดยใช้ข้อสนเทศของแฮนแนนและควินน์ (HQ)

5. คำนวณค่าอัตราความถูกต้อง (%) ของการคัดเลือกจำนวนปัจจัยที่ถูกต้องของเกณฑ์ทั้ง 4 เกณฑ์ เมื่อ $m = \hat{m}$ จึงจัดว่าการคัดเลือกจำนวนปัจจัยถูกต้อง โดยการทำซ้ำ 400 รอบ

6. ทำการเปรียบเทียบประสิทธิภาพการคัดเลือกจำนวนปัจจัยของทั้ง 4 เกณฑ์ โดยพิจารณาจากอัตราความถูกต้องที่ได้จากข้อ 5 และกราฟระหว่างสัดส่วนจำนวนปัจจัยต่อจำนวนตัวแปรและอัตราความถูกต้องของเกณฑ์ทั้ง 4 เกณฑ์ แล้วยืนยันผลที่ได้ด้วยการทดสอบสมมติฐานเปรียบเทียบผลต่างระหว่างค่าสัดส่วนความถูกต้องของเกณฑ์การคัดเลือกจำนวนปัจจัยแต่ละคู่

7. สรุปผลการวิจัย

โดยแผนผังแสดงขั้นตอนในการดำเนินการวิจัย ดังแสดงในรูปที่ 3.1



รูปที่ 3.1 แสดงแผนผังขั้นตอนในการดำเนินการวิจัย

สำหรับในแต่ละขั้นตอนมีรายละเอียดดังนำเสนอเป็นลำดับต่อไปนี้

ขั้นตอนที่ 1

กำหนดลักษณะของข้อมูลตามจำนวนตัวแปร (p) จำนวนปัจจัย (m) และขนาดตัวอย่าง (n) ที่กำหนดในขอบเขตการวิจัย

ขั้นตอนที่ 2

สร้างเมทริกซ์สหสัมพันธ์ Σ_p ด้วยวิธี Onion เพื่อสร้างเมทริกซ์ความแปรปรวนร่วม (Covariance Matrix : Σ) ที่มีค่าแปรปรวน (Variance) เท่ากับ 1 สำหรับตัวแปร X จำนวน p ตัวแปร ซึ่งจะทำให้ได้ Factor Loading สำหรับ m ปัจจัย และเมทริกซ์ความแปรปรวนของค่าเฉพาะ (ψ) ซึ่งกำหนดให้ p = 10, 20, 30 และ 40 โดยมีขั้นตอนดังนี้

1. สร้างเมทริกซ์สหสัมพันธ์ที่เป็นไปได้ (Feasible Correlation Matrix) Σ_p ด้วยวิธี Onion ซึ่งรายละเอียดและขั้นตอนได้กล่าวถึงในบทที่ 2 หัวข้อที่ 1

2. แยกเมทริกซ์สหสัมพันธ์ Σ_p ที่ได้จากข้อ 1 ด้วย Singular Value Decomposition (SVD) เพื่อหาเมทริกซ์ของ Factor Loading ขนาด pxm และเมทริกซ์ความแปรปรวนของค่าเฉพาะ (ψ) ขนาด pxp โดยมีรายละเอียดดังนี้

ทำการแยกเมทริกซ์สหสัมพันธ์ Σ_p เพื่อหาเมทริกซ์ $\tilde{D}$ และ $\tilde{V}$ ด้วย SVD นั่นคือ

$$\Sigma_p = \tilde{V}\tilde{D}\tilde{V}^t$$

เมื่อ $\tilde{D}$ คือ เมทริกซ์ทแยงมุม (Diagonal Matrix) ขนาด p x p ($D \in R^{pxp}$)

$\tilde{V}$ คือ เมทริกซ์ที่หลักถูกปรับปกติขนาด p x p ($V \in R^{pxp}$)

ต่อไปทำการจัด $\Sigma_p = \tilde{V}\tilde{D}\tilde{V}^t$ ให้อยู่ในรูปของ Factor Loading และ ψ ที่มี m ปัจจัย ได้

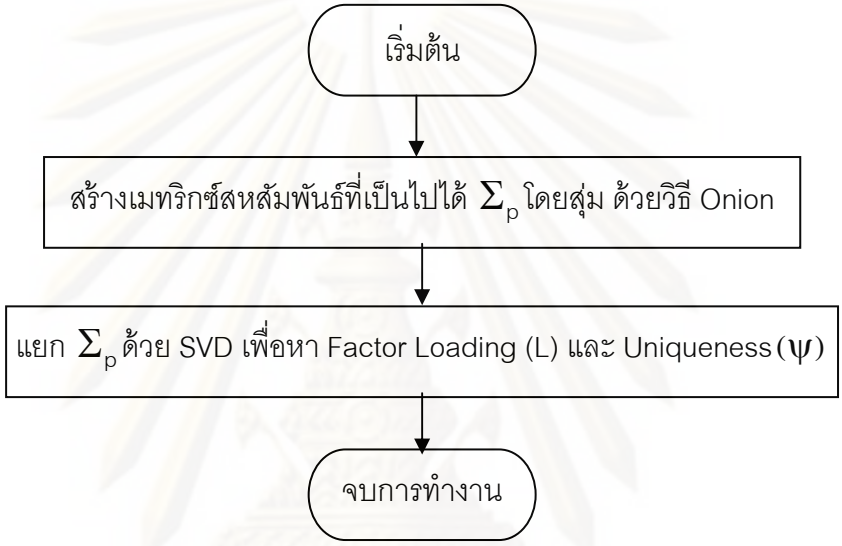
$$\Sigma = VDV^t + \psi$$

เมื่อ D คือ เมทริกซ์ทแยงมุมขนาด m x m ($D \in R^{mxm}$) โดยที่ ค่าบนเส้นทแยงมุม คือ m ค่าแรกบนเส้นทแยงมุมของเมทริกซ์ $\tilde{D}$

V คือ เมทริกซ์ที่หลักถูกปรับปกติขนาด p x m ($V \in R^{pxm}$) โดยที่ ประกอบด้วย m หลักแรกของเมทริกซ์ $\tilde{V}$

ดังนั้น Factor Loading (L) $= VD^{1/2}$

และ ψ คือ เมทริกซ์ทแยงมุมขนาด p x p ($\psi \in R^{pxp}$) โดยที่ $\psi_i = \Sigma_{p_{ii}} - \sum_{j=1}^{m} \ell_{ij}^2$ เมื่อ ℓ_{ij} คือ Factor Loading ของตัวแปรที่ i บนปัจจัยที่ j สำหรับขั้นตอนการสร้าง Factor Loading และ ψ สามารถเขียนแผนผังได้ดังนี้



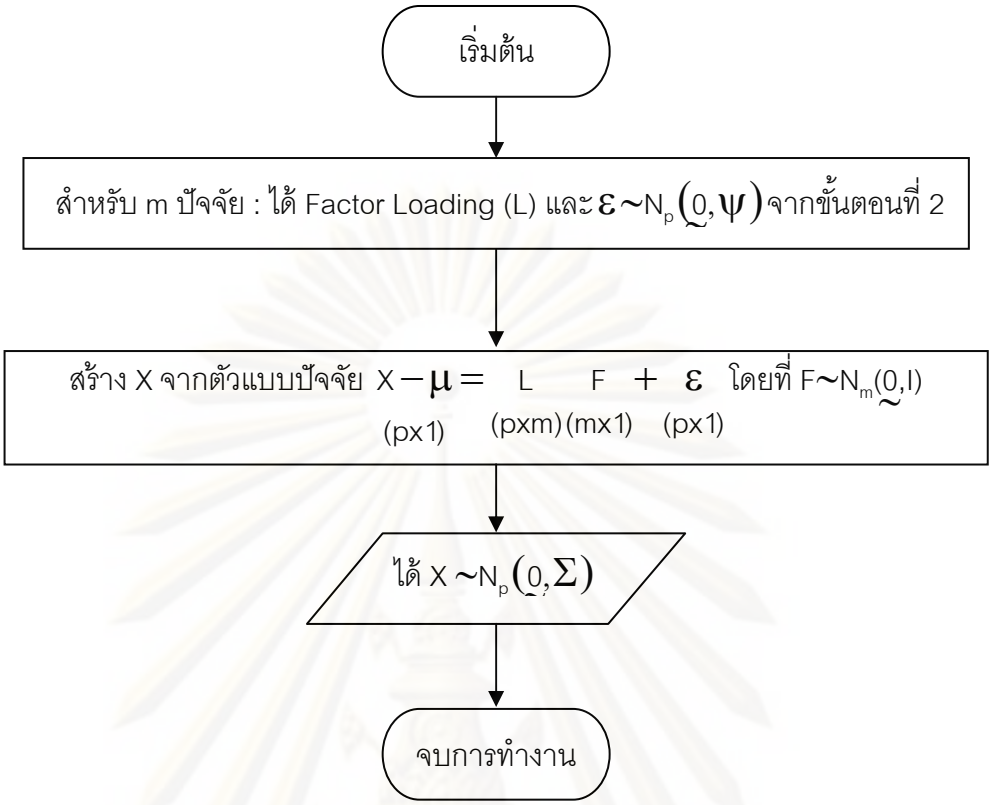
รูปที่ 3.2 แสดงแผนผังขั้นตอนการสร้าง factor loading และ ψ

ขั้นตอนที่ 3

สร้างข้อมูลของตัวแปร X จำนวน p ตัวแปร ที่มีการแจกแจงแบบปกติหลายตัวแปร (Multivariate Normal Distribution) มีเวกเตอร์ค่าเฉลี่ยเท่ากับ $\underset{\sim}{0}$ และเมทริกซ์ความแปรปรวนร่วม Σ ที่ได้จากขั้นตอนที่ 2 ($X \sim N_p(\underset{\sim}{0}, \Sigma)$) จากตัวแบบปัจจัย คือ

$$\underset{(nxp)}{X - \mu} = \underset{(nxm)}{L} \quad \underset{(mxp)}{F} + \underset{(nxp)}{\varepsilon}$$

โดยมี p, m และ n ตามที่กำหนดในขั้นตอนที่ 1 Factor Loading (L) และ $\varepsilon \sim N_p(\underset{\sim}{0}, \psi)$ ได้จากขั้นตอนที่ 2 และ Factor Score (F) $\sim N_m(\underset{\sim}{0}, I)$ สำหรับขั้นตอนการสร้างข้อมูล X สามารถเขียนแผนผังได้ดังนี้



รูปที่ 3.3 แสดงแผนผังขั้นตอนการสร้างข้อมูล X

ขั้นตอนที่ 4

สมมติว่าไม่ทราบค่า m ทำการคัดเลือกจำนวนปัจจัยที่เหมาะสม ($\hat{m}$) ด้วยเกณฑ์ 4 เกณฑ์ ได้แก่

4.1 10-fold Likelihood Cross-Validation (LCV)

4.2 เกณฑ์การคัดเลือกจำนวนปัจจัยโดยใช้ข้อสนเทศของอากาอิเกะ (AIC)

4.3 เกณฑ์การคัดเลือกจำนวนปัจจัยโดยใช้ข้อสนเทศของชวาร์ซ (SIC)

4.4 เกณฑ์การคัดเลือกจำนวนปัจจัยโดยใช้ข้อสนเทศของแฮนแนนและควินน์ (HQ)

วิธีการคัดเลือกจำนวนปัจจัยที่เหมาะสมของเกณฑ์การคัดเลือกจำนวนปัจจัยแต่ละเกณฑ์มีขั้นตอนดังนี้

4.1 10-fold Likelihood Cross-Validation (LCV)

4.1.1 กำหนด n, p และ m = 1, 2, … , p

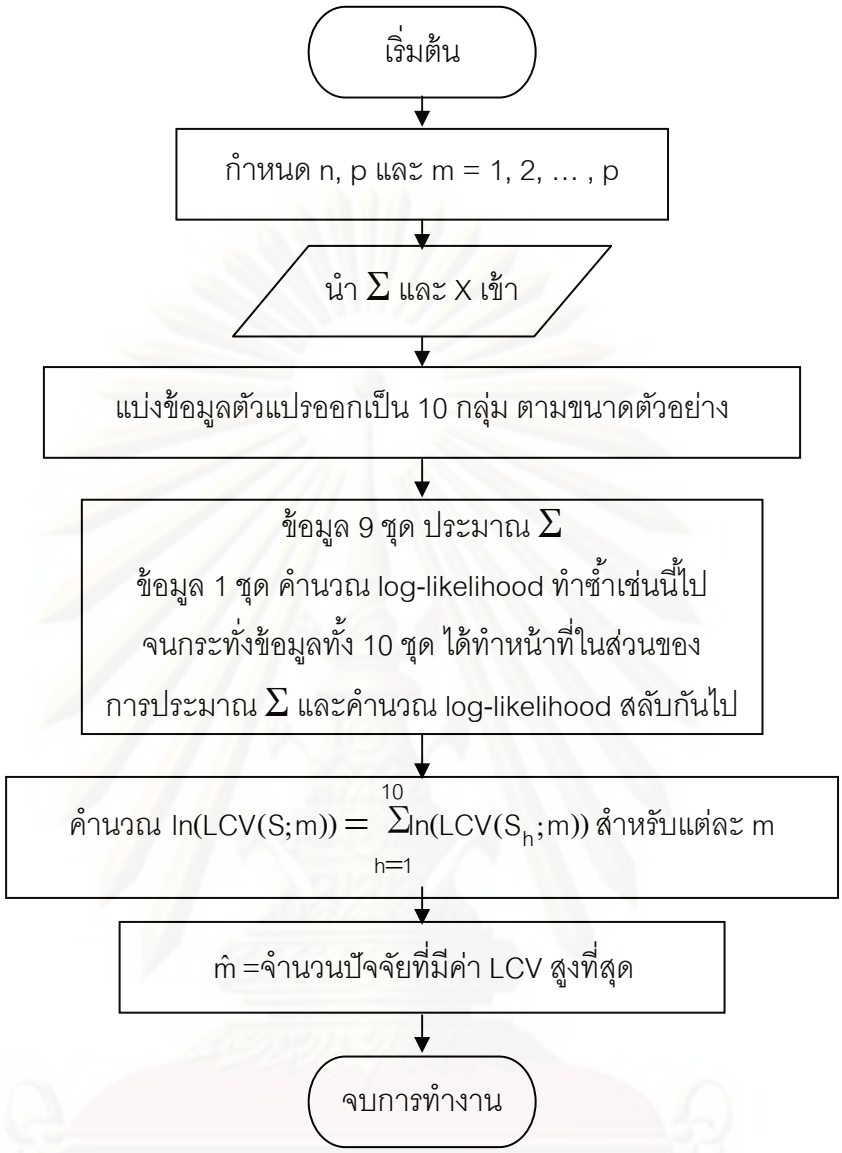
4.1.2 นำ Σ และ X เข้า

4.1.3 แบ่งข้อมูลตัวแปร X ออกเป็น 10 กลุ่ม ตามขนาดตัวอย่าง (n)

4.1.4 นำข้อมูล 9 ชุด ประมาณค่าเมทริกซ์ความแปรปรวนร่วม Σ และข้อมูลอีก 1 ชุดที่เหลือคำนวณค่า log-likelihood ดังสมการที่ (14) ซึ่งจะคำนวณในขั้นตอนนี้ทั้งหมด 10 รอบ จนกระทั่งข้อมูลทั้ง 10 ชุด ได้ทำหน้าที่ในส่วนของการประมาณเมทริกซ์ความแปรปรวนร่วมและคำนวณค่า log-likelihood สลับกันไป ดังรูปที่ 2.3 แสดงขั้นตอนการทำงานของ 10-fold Likelihood Cross-Validation

4.1.5 คำนวณค่า log-likelihood สำหรับแต่ละจำนวนปัจจัยดังสมการ (16)

4.1.6 ทำการเปรียบเทียบค่า LCV ของทุกจำนวนปัจจัย (m) โดย $\hat{m}$ เท่ากับจำนวนปัจจัยที่ให้ค่า LCV สูงที่สุด



รูปที่ 3.4 แสดงแผนผังขั้นตอนการคัดเลือกจำนวนปัจจัยของเกณฑ์ LCV

4.2 เกณฑ์การคัดเลือกจำนวนปัจจัยโดยใช้ข้อสนเทศของอากาอิเกะ (AIC)

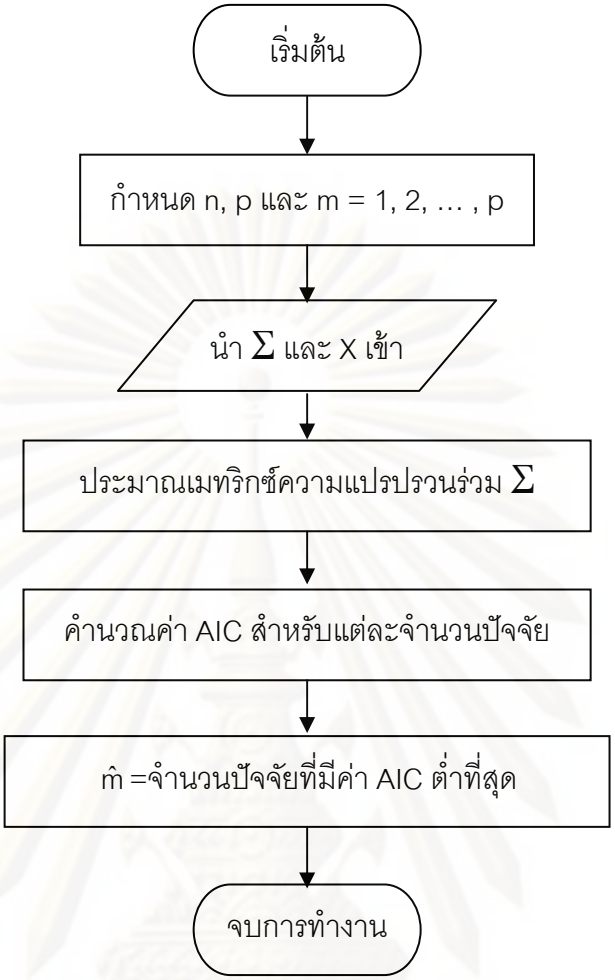
4.2.1 กำหนด n, p และ m = 1, 2, … , p

4.2.2 นำ Σ และ X เข้า

4.2.3 ประมาณค่าเมทริกซ์ความแปรปรวนร่วม Σ

4.2.4 คำนวณค่า AIC สำหรับแต่ละจำนวนปัจจัย

4.2.5 ทำการเปรียบเทียบค่า AIC ของทุกจำนวนปัจจัย (m) ซึ่ง $\hat{m}$ เท่ากับจำนวนปัจจัยที่ให้ค่า AIC ต่ำที่สุด



รูปที่ 3.5 แสดงแผนผังขั้นตอนการคัดเลือกจำนวนปัจจัยของเกณฑ์ AIC

4.3 เกณฑ์การคัดเลือกจำนวนปัจจัยโดยใช้ข้อสนเทศของชวาร์ช (SIC)

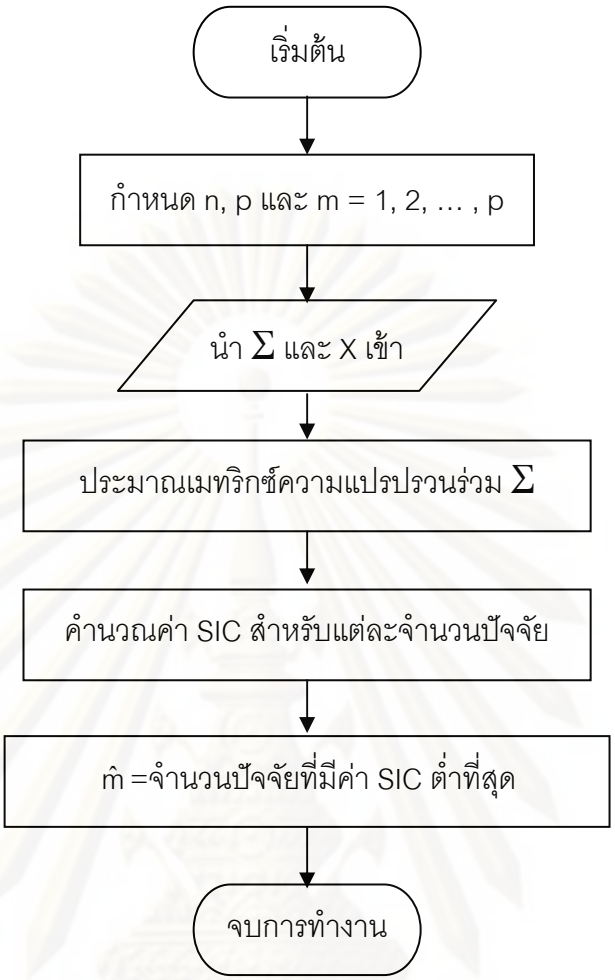
4.3.1 4.3.1กำหนด n, p และ m = 1, 2, … , p

4.3.2 4.3.2นำ Σ และ X เข้า

4.3.3 ประมาณค่าเมทริกซ์ความแปรปรวนร่วม Σ

4.3.4 คำนวณค่า SIC สำหรับแต่ละจำนวนปัจจัย

4.3.5 ทำการเปรียบเทียบค่า SIC ของทุกจำนวนปัจจัย (m) ซึ่ง $\hat{m}$ เท่ากับจำนวนปัจจัยที่ให้ค่า SIC ต่ำที่สุด



รูปที่ 3.6 แสดงแผนผังขั้นตอนการคัดเลือกจำนวนปัจจัยของเกณฑ์ SIC

4. **เกณฑ์การคัดเลือกจำนวนปัจจัยโดยใช้ข้อสนเทศของแฮนแนนและควินน์** (HQ)

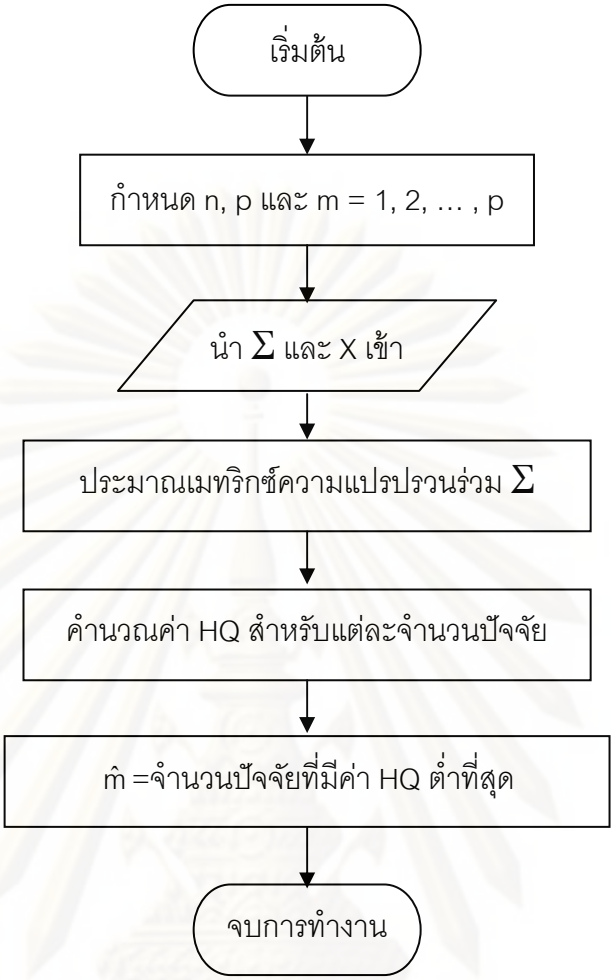
4.4.1 กำหนด n, p และ m = 1, 2, … , p

4.4.2 นำ Σ และ X เข้า

4.4.3 ประมาณค่าเมทริกซ์ความแปรปรวนร่วม Σ

4.4.4 คำนวณค่า HQ สำหรับแต่ละจำนวนปัจจัย

4.4.5 ทำการเปรียบเทียบค่า HQ ของทุกจำนวนปัจจัย (m) ซึ่ง $\hat{m}$ เท่ากับจำนวนปัจจัยที่ให้ค่า HQ ต่ำที่สุด



รูปที่ 3.7 แสดงแผนผังขั้นตอนการคัดเลือกจำนวนปัจจัยของเกณฑ์ HQ

ขั้นตอนที่ 5

ทำการคำนวณค่าอัตราความถูกต้อง (%) ของการคัดเลือกจำนวนปัจจัยที่ถูกต้องของเกณฑ์ทั้ง 4 เกณฑ์จากการทำซ้ำ 400 รอบ โดยที่ความถูกต้องในการคัดเลือกจำนวนปัจจัย คือ $m = \hat{m}$ ดังนั้น

$$\text{อัตราความถูกต้อง(\%)} = \frac{\text{จำนวนรอบที่คัดเลือกจำนวนปัจจัยถูกต้อง}(m=\hat{m})}{400} \times 100$$

เป็นเกณฑ์ที่ใช้เปรียบเทียบประสิทธิภาพของทั้ง 4 เกณฑ์ข้างต้น

ขั้นตอนที่ 6

ทำการเปรียบเทียบประสิทธิภาพการคัดเลือกจำนวนปัจจัยของทั้ง 4 เกณฑ์ โดยพิจารณาจากอัตราความถูกต้องที่ได้จากขั้นตอนที่ 5 และกราฟระหว่างสัดส่วนจำนวนปัจจัยต่อจำนวนตัวแปร (%) และอัตราความถูกต้อง (%) ของเกณฑ์ทั้ง 4 เกณฑ์ โดยเกณฑ์ที่มีอัตราความถูกต้องมากที่สุดเป็นเกณฑ์ที่มีประสิทธิภาพสูงที่สุด และยืนยันผลที่ได้ด้วยการทดสอบสมมติฐานเปรียบเทียบผลต่างระหว่างค่าสัดส่วนความถูกต้องของเกณฑ์การคัดเลือกจำนวนปัจจัยแต่ละคู่

ขั้นตอนที่ 7

สรุปผลการวิจัยทั้งหมดในรูปของตารางและรูปภาพเพื่อแสดงการเปรียบเทียบประสิทธิภาพของเกณฑ์ทั้ง 4 เกณฑ์

บทที่ 4

ผลการวิเคราะห์ข้อมูล

งานวิจัยนี้เป็นการศึกษาการเปรียบเทียบเกณฑ์การคัดเลือกจำนวนปัจจัยในการวิเคราะห์ปัจจัยเชิงสถิติ ซึ่งเกณฑ์การคัดเลือกจำนวนปัจจัยที่นำมาพิจารณามี 4 เกณฑ์ ดังนี้

1. 10-fold Likelihood Cross-Validation (LCV)
2. เกณฑ์การคัดเลือกจำนวนปัจจัยโดยใช้ข้อสนเทศของอากาอิเกะ (Akaike's Information Criteria : AIC)
3. เกณฑ์การคัดเลือกจำนวนปัจจัยโดยใช้ขอสนเทศของชวาร์ซ (Swcharz's Information Criteria : SIC)
4. เกณฑ์การคัดเลือกจำนวนปัจจัยโดยใช้ข้อสนเทศของแฮนแนนและควินน์ (Hannan and Quinn's Information Criteria : HQ)

เกณฑ์ที่ใช้ในการตัดสินใจว่า เกณฑ์การคัดเลือกจำนวนปัจจัยเกณฑ์ใดมีประสิทธิภาพสูงที่สุด จะพิจารณาจากอัตราความถูกต้อง (%) ซึ่งเกณฑ์การคัดเลือกจำนวนปัจจัยที่มีอัตราความถูกต้องมากที่สุด จะเป็นเกณฑ์ที่มีประสิทธิภาพสูงที่สุด

ขอบเขตของการวิจัยจะเป็นการศึกษาตัวแบบปัจจัย (Factor Model) ตามสถานการณ์ที่กำหนดดังตารางต่อไปนี้

สถานการณ์ที่	จำนวนตัวแปร (p)	จำนวนปัจจัย (m)	ขนาดตัวอย่าง (n)
1	10	1,2, …, 5	200, 300, 500 และ 1,000
2	20	1,2, …, 10	300, 500 และ 1,000
3	30	1,2, …, 15	500 และ 1,000
4	40	1,2, …, 20	1,000

การคัดเลือกจำนวนปัจจัยจะใช้ในกรณีที่ตัวแปรมีการแจกแจงแบบปกติหลายตัวแปรที่มีเวกเตอร์ค่าเฉลี่ย $\underset{\sim}{0}$ และเมทริกซ์ความแปรปรวนร่วม Σ โดยที่มีค่าแปรปรวนเท่ากับ 1 และทำการคัดเลือกจำนวนปัจจัยสำหรับเกณฑ์แต่ละเกณฑ์ทั้งหมด 400 รอบในแต่ละสถานการณ์

สำหรับการนำเสนอผลการวิจัย ผู้วิจัยได้กำหนดสัญลักษณ์ต่างๆ เพื่อใช้ในตารางและการสรุปผลดังนี้

n แทน ขนาดตัวอย่าง

m แทน จำนวนปัจจัย

LCV แทน 10-fold Likelihood Cross-Validation

AIC แทน เกณฑ์การคัดเลือกจำนวนปัจจัยโดยใช้ข้อสนเทศของอากาอิเกะ

SIC แทน เกณฑ์การคัดเลือกจำนวนปัจจัยโดยใช้ข้อสนเทศของชวาร์ซ

HQ แทน เกณฑ์การคัดเลือกจำนวนปัจจัยโดยใช้ข้อสนเทศของแฮนแนนและควินน์

p_{LCV} แทน ค่าสัดส่วนของอัตราความถูกต้องของการคัดเลือกจำนวนปัจจัยโดยใช้เกณฑ์ LCV

p_{AIC} แทน ค่าสัดส่วนของอัตราความถูกต้องของการคัดเลือกจำนวนปัจจัยโดยใช้เกณฑ์ AIC

p_{SIC} แทน ค่าสัดส่วนของอัตราความถูกต้องของการคัดเลือกจำนวนปัจจัยโดยใช้เกณฑ์ SIC

p_{HQ} แทน ค่าสัดส่วนของอัตราความถูกต้องของการคัดเลือกจำนวนปัจจัยโดยใช้เกณฑ์ HQ

การนำเสนอผลการวิจัยจะนำเสนอด้วยอัตราความถูกต้อง (%) ของการคัดเลือกจำนวนปัจจัยที่ถูกต้องของเกณฑ์แต่ละเกณฑ์โดยแสดงด้วยตัวเลขที่มีทศนิยม 2 ตำแหน่ง กราฟระหว่างสัดส่วนของจำนวนปัจจัยต่อจำนวนตัวแปร (%) และการทดสอบสมมติฐานเปรียบเทียบค่าสัดส่วนความถูกต้องของการคัดเลือกจำนวนปัจจัยทั้ง 4 เกณฑ์ ด้วยค่า p-value โดยแสดงด้วยตัวเลขทศนิยม 4 ตำแหน่ง ตามสถานการณ์ต่างๆ ที่กำหนด โดยแบ่งการนำเสนอเป็น 4 ส่วน ดังนี้

ตอนที่ 4.1 ผลการวิจัยของการเปรียบเทียบเกณฑ์การคัดเลือกจำนวนปัจจัยในการวิเคราะห์ปัจจัย กรณีจำนวนตัวแปรเท่ากับ 10

ตอนที่ 4.2 ผลการวิจัยของการเปรียบเทียบเกณฑ์การคัดเลือกจำนวนปัจจัยในการวิเคราะห์ปัจจัย กรณีจำนวนตัวแปรเท่ากับ 20

ตอนที่ 4.3 ผลการวิจัยของการเปรียบเทียบเกณฑ์การคัดเลือกจำนวนปัจจัยในการวิเคราะห์ปัจจัย กรณีจำนวนตัวแปรเท่ากับ 30

ตอนที่ 4.4 ผลการวิจัยของการเปรียบเทียบเกณฑ์การคัดเลือกจำนวนปัจจัยในการวิเคราะห์ปัจจัย กรณีจำนวนตัวแปรเท่ากับ 40

ตอนที่ 4.1 ผลการวิจัยของการเปรียบเทียบเกณฑ์การคัดเลือกจำนวนปัจจัยในการวิเคราะห์ปัจจัย กรณีจำนวนตัวแปรเท่ากับ 10

4.1.1 พิจารณาอัตราความถูกต้องของการคัดเลือกจำนวนปัจจัย ดังแสดงในตารางที่ 4.1-4.4 และรูปที่ 4.1-4.4

ตารางที่ 4.1 แสดงค่าอัตราความถูกต้องของเกณฑ์ LCV, เกณฑ์ AIC, เกณฑ์ SIC และเกณฑ์ HQ ในการวิเคราะห์ปัจจัย เมื่อจำนวนตัวแปรเท่ากับ 10 จำนวนปัจจัยเท่ากับ 1, 2, ..., 5 และขนาดตัวอย่างเท่ากับ 200

criteria / m	LCV	AIC	SIC	HQ
1	100.00	100.00	100.00	100.00
2	100.00	100.00	100.00	100.00
3	100.00	99.25	100.00	100.00
4	93.75	80.75	98.25	91.75
5	67.50	58.25	75.25	65.25

หมายเหตุ ค่าอัตราความถูกต้องเป็นค่าจากการทำการวิเคราะห์ปัจจัยทั้งหมด 400 รอบ

ตารางที่ 4.2 แสดงค่าอัตราความถูกต้องของเกณฑ์ LCV, เกณฑ์ AIC, เกณฑ์ SIC และเกณฑ์ HQ ในการวิเคราะห์ปัจจัย เมื่อจำนวนตัวแปรเท่ากับ 10 จำนวนปัจจัยเท่ากับ 1, 2, ..., 5 และขนาดตัวอย่างเท่ากับ 300

criteria / m	LCV	AIC	SIC	HQ
1	100.00	100.00	100.00	100.00
2	100.00	100.00	100.00	100.00
3	99.00	98.00	100.00	99.75
4	88.25	74.00	94.75	82.25
5	68.75	57.00	71.00	59.75

หมายเหตุ ค่าอัตราความถูกต้องเป็นค่าจากการทำการวิเคราะห์ปัจจัยทั้งหมด 400 รอบ

ตารางที่ 4.3 แสดงค่าอัตราความถูกต้องของเกณฑ์ LCV, เกณฑ์ AIC, เกณฑ์ SIC และเกณฑ์ HQ ในการวิเคราะห์ปัจจัย เมื่อจำนวนตัวแปรเท่ากับ 10 จำนวนปัจจัยเท่ากับ 1, 2, ..., 5 และขนาดตัวอย่างเท่ากับ 500

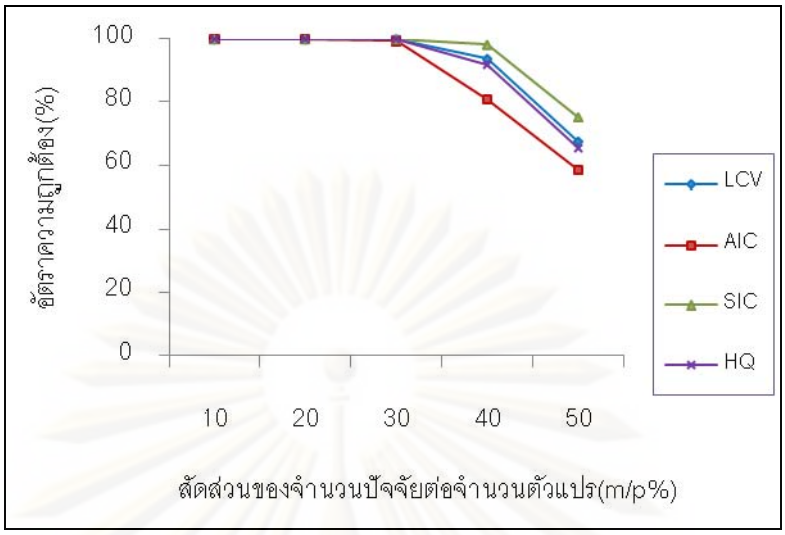
criteria / m	LCV	AIC	SIC	HQ
1	100.00	100.00	100.00	100.00
2	100.00	99.75	100.00	100.00
3	95.00	92.25	99.75	98.50
4	78.00	70.75	82.75	77.50
5	54.25	48.25	54.50	55.75

หมายเหตุ ค่าอัตราความถูกต้องเป็นค่าจากการทำการวิเคราะห์ปัจจัยทั้งหมด 400 รอบ

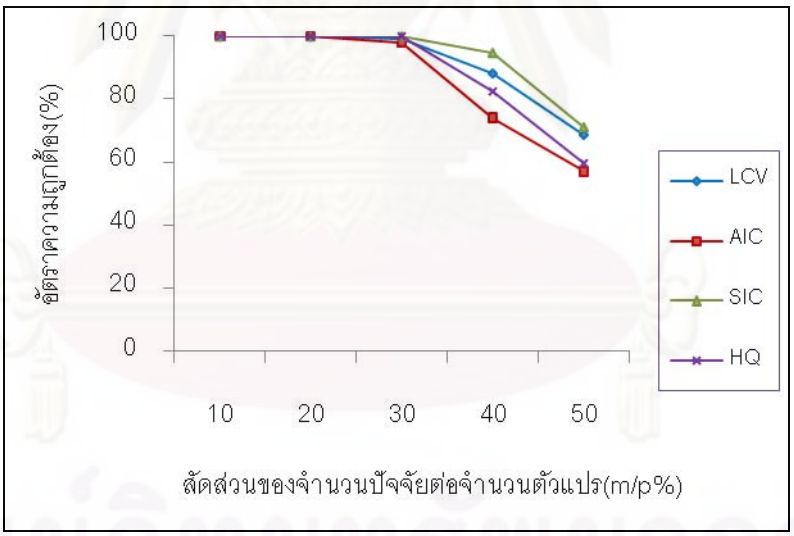
ตารางที่ 4.4 แสดงค่าอัตราความถูกต้องของเกณฑ์ LCV, เกณฑ์ AIC, เกณฑ์ SIC และเกณฑ์ HQ ในการวิเคราะห์ปัจจัย เมื่อจำนวนตัวแปรเท่ากับ 10 จำนวนปัจจัยเท่ากับ 1, 2, ..., 5 และขนาดตัวอย่างเท่ากับ 1,000

criteria / m	LCV	AIC	SIC	HQ
1	100.00	100.00	100.00	100.00
2	100.00	100.00	100.00	100.00
3	91.25	87.50	98.50	93.00
4	68.50	66.50	76.50	66.75
5	49.25	51.00	54.25	49.25

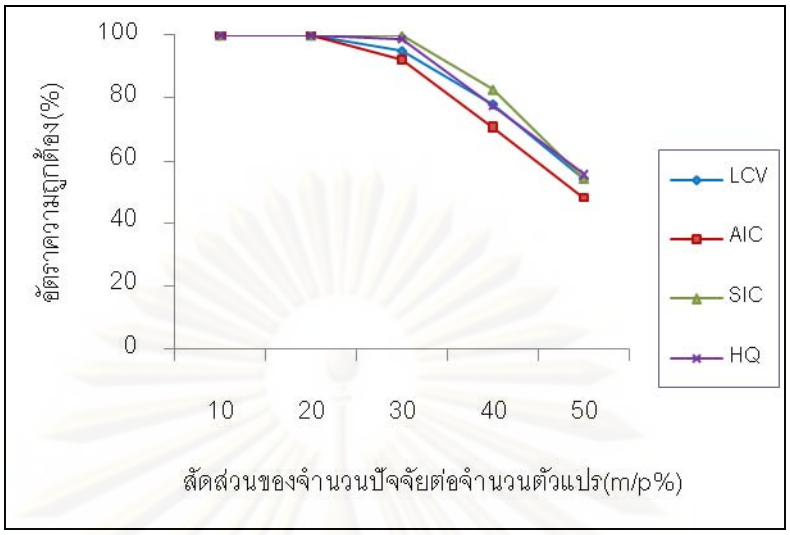
หมายเหตุ ค่าอัตราความถูกต้องเป็นค่าจากการทำการวิเคราะห์ปัจจัยทั้งหมด 400 รอบ



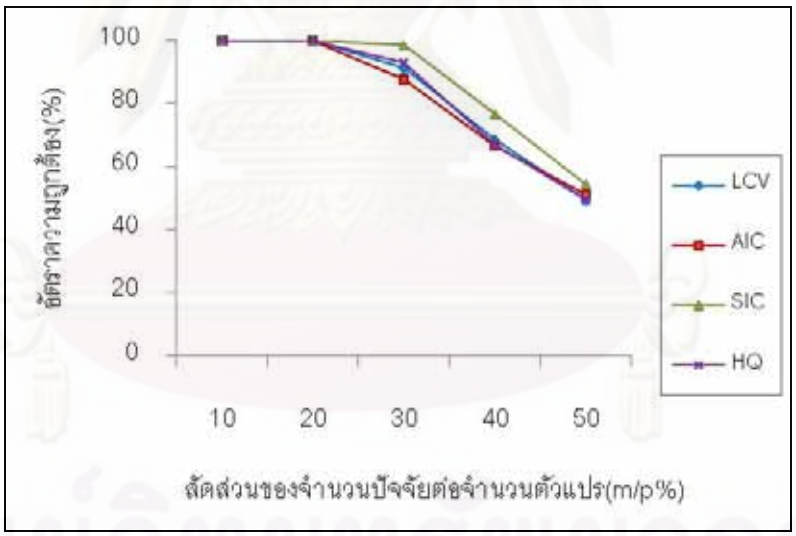
รูปที่ 4.1 แสดงการเปรียบเทียบเกณฑ์การคัดเลือกจำนวนปัจจัยในการวิเคราะห์ปัจจัยด้วยอัตราความถูกต้อง (%) สำหรับจำนวนตัวแปรเท่ากับ 10 และขนาดตัวอย่างเท่ากับ 200



รูปที่ 4.2 แสดงการเปรียบเทียบเกณฑ์การคัดเลือกจำนวนปัจจัยในการวิเคราะห์ปัจจัยด้วยอัตราความถูกต้อง(%) สำหรับจำนวนตัวแปรเท่ากับ 10 และขนาดตัวอย่างเท่ากับ 300



รูปที่ 4.3 แสดงการเปรียบเทียบเกณฑ์การคัดเลือกจำนวนปัจจัยในการวิเคราะห์ปัจจัย ด้วยอัตราความถูกต้อง(%) สำหรับจำนวนตัวแปรเท่ากับ 10 และขนาดตัวอย่างเท่ากับ 500



รูปที่ 4.4 แสดงการเปรียบเทียบเกณฑ์การคัดเลือกจำนวนปัจจัยในการวิเคราะห์ปัจจัย ด้วยอัตราความถูกต้อง(%) สำหรับจำนวนตัวแปรเท่ากับ 10 และขนาดตัวอย่างเท่ากับ 1,000

4.1.2 **พิจารณาผลการทดสอบสมมติฐานของค่าสัดส่วนความถูกต้องของการคัดเลือกจำนวนปัจจัย ดังแสดงในตารางที่** 4.5

ตารางต่อไปนี้เป็นตารางแสดงค่า p-value สำหรับการทดสอบสมมติฐานเปรียบเทียบค่าสัดส่วนความถูกต้องของการคัดเลือกจำนวนปัจจัยในการวิเคราะห์ปัจจัย ซึ่งได้แสดงค่า p-value ไว้เพียงบางค่าเท่านั้น โดยพิจารณาจากกราฟข้างต้นแล้วพบว่า เกณฑ์การคัดเลือกจำนวนปัจจัยบางเกณฑ์มีค่าสัดส่วนความถูกต้องแตกต่างกันไม่ชัดเจน จึงได้ทำการทดสอบสมมติฐานเพื่อยืนยันผลที่ได้

ตารางที่ 4.5 แสดงค่า p-value ของการทดสอบสมมติฐานเปรียบเทียบค่าสัดส่วนความถูกต้องของการคัดเลือกจำนวนปัจจัยในการวิเคราะห์ปัจจัย กรณีจำนวนตัวแปรเท่ากับ 10 ที่ระดับนัยสำคัญ 0.05

n	m	$H_0:p_{LCV} \leq p_{AIC}$ $H_1:p_{LCV} > p_{AIC}$	$H_0:p_{SIC} \leq p_{LCV}$ $H_1:p_{SIC} > p_{LCV}$	$H_0:p_{LCV} \leq p_{HQ}$ $H_1:p_{LCV} > p_{HQ}$	$H_0:p_{SIC} \leq p_{AIC}$ $H_1:p_{SIC} > p_{AIC}$	$H_0:p_{HQ} \leq p_{AIC}$ $H_1:p_{HQ} > p_{AIC}$	$H_0:p_{SIC} \leq p_{HQ}$ $H_1:p_{SIC} > p_{HQ}$
300	3	0.1223	0.0225**	0.9108	0.0022**	0.0095**	0.1585
	5	0.0003**	0.2440	0.0040**	0.0000**	0.2151	0.0004**
500	3	0.0557	0.0000**	0.9974	0.0000**	0.0000**	0.0288**
	5	0.0448**	0.4717	0.6651	0.0385**	0.0169**	0.6389
1,000	5	0.6897	0.0785	0.5000	0.1787	0.6897	0.0785

หมายเหตุ 1. ** หมายถึง การปฏิเสธสมมติฐานหลักที่ระดับนัยสำคัญ 0.05

2. ตารางนี้แสดงค่า p-value ไว้เพียงบางส่วนเท่านั้น สามารถพิจารณาค่า p-value ทั้งหมดได้ในส่วนของภาคผนวก ข

จากผลการวิจัยของการเปรียบเทียบประสิทธิภาพเกณฑ์การคัดเลือกจำนวนปัจจัยในการวิเคราะห์ปัจจัยสำหรับตัวแบบปัจจัยที่ประกอบด้วยตัวแปร 10 ตัวแปร และจำนวนปัจจัยเท่ากับ 1, 2, …, 5 (ตารางที่ 4.1-4.5 และรูปที่ 4.1-4.4) สามารถอธิบายรายละเอียดได้ดังนี้

ขนาดตัวอย่างเท่ากับ 200

สำหรับจำนวนปัจจัยไม่เกินร้อยละ 20 ของจำนวนตัวแปร เกณฑ์การคัดเลือกจำนวนปัจจัยทั้ง 4 เกณฑ์ มีอัตราความถูกต้องสูงเท่ากัน ที่จำนวนปัจจัยมากกว่าร้อยละ 20 แต่ไม่เกินร้อยละ 30 ของจำนวนตัวแปร เกณฑ์ LCV เกณฑ์ HQ และเกณฑ์ SIC ยังมีอัตราความถูกต้องเท่ากัน แต่มากกว่าเกณฑ์ AIC และเมื่อจำนวนปัจจัยเพิ่มขึ้นมากกว่าร้อยละ 30 ของจำนวนตัวแปร พบว่า เกณฑ์ SIC มีอัตราความถูกต้องสูงที่สุด รองลงมา คือ เกณฑ์ LCV และเกณฑ์ HQ ซึ่งมีอัตราความถูกต้องไม่แตกต่างกัน และเกณฑ์ AIC มีอัตราความถูกต้องต่ำที่สุด

ขนาดตัวอย่างเท่ากับ 300

สำหรับจำนวนปัจจัยไม่เกินร้อยละ 20 ของจำนวนตัวแปร เกณฑ์การคัดเลือกจำนวนปัจจัยทั้ง 4 เกณฑ์ มีอัตราความถูกต้องสูงเท่ากัน ที่จำนวนปัจจัยมากกว่าร้อยละ 20 แต่ไม่เกินร้อยละ 30 ของจำนวนตัวแปร เกณฑ์ SIC และเกณฑ์ HQ มีอัตราความถูกต้องไม่แตกต่างกัน แต่มากกว่าเกณฑ์ LCV และเกณฑ์ AIC มีอัตราความถูกต้องต่ำที่สุด เมื่อจำนวนปัจจัยเพิ่มขึ้นเป็นร้อยละ 40 ของจำนวนตัวแปร พบว่า เกณฑ์ SIC มีอัตราความถูกต้องสูงที่สุด รองลงมา คือ เกณฑ์ LCV เกณฑ์ HQ และเกณฑ์ AIC ตามลำดับ แต่เมื่อจำนวนปัจจัยเท่ากับร้อยละ 50 ของจำนวนตัวแปร เกณฑ์ SIC และเกณฑ์LCV มีอัตราความถูกต้องไม่แตกต่างกันและมากกว่าเกณฑ์ HQ และเกณฑ์ AIC ซึ่งมีอัตราความถูกต้องไม่แตกต่างกัน

ขนาดตัวอย่างเท่ากับ 500

สำหรับจำนวนปัจจัยไม่เกินร้อยละ 20 ของจำนวนตัวแปร เกณฑ์การคัดเลือกจำนวนปัจจัยทั้ง 4 เกณฑ์ มีอัตราความถูกต้องสูงเท่ากัน ที่จำนวนปัจจัยมากกว่าร้อยละ 20 แต่ไม่เกินร้อยละ 40 ของจำนวนตัวแปร เกณฑ์ SIC ยังคงมีอัตราความถูกต้องสูงที่สุด รองลงมา คือ เกณฑ์ HQ ซึ่งมีอัตราความถูกต้องไม่แตกต่างจากเกณฑ์ LCV แต่มากกว่าเกณฑ์ AIC และเมื่อจำนวนปัจจัยมากกว่าร้อยละ 40 ของจำนวนตัวแปร พบว่า เกณฑ์ทั้ง 4 เกณฑ์มีอัตราความถูกต้องลดลงจนกระทั่งเกณฑ์ SIC เกณฑ์ LCV และเกณฑ์ HQ มีอัตราความถูกต้องไม่แตกต่างกันแต่เกณฑ์ AIC ยังคงมีอัตราความถูกต้องต่ำที่สุด

ขนาดตัวอย่างเท่ากับ 1,000

สำหรับจำนวนปัจจัยไม่เกินร้อยละ 20 ของจำนวนตัวแปร เกณฑ์การคัดเลือกจำนวนปัจจัยทั้ง 4 เกณฑ์ มีอัตราความถูกต้องสูงเท่ากัน ที่จำนวนปัจจัยมากกว่าร้อยละ 20 แต่ไม่

เกินร้อยละ 40 ของจำนวนตัวแปร เกณฑ์ SIC มีอัตราความถูกต้องสูงที่สุด รองลงมา คือ เกณฑ์ LCV และเกณฑ์ HQ ซึ่งมีอัตราความถูกต้องไม่แตกต่างกัน และเกณฑ์ AIC มีอัตราความถูกต้องต่ำที่สุด และเมื่อจำนวนปัจจัยมากกว่าร้อยละ 40 ของจำนวนตัวแปร พบว่า เกณฑ์ทั้ง 4 เกณฑ์ มีอัตราความถูกต้องไม่แตกต่างกัน

เมื่อพิจารณาอัตราความถูกต้องของเกณฑ์ทั้ง 4 เกณฑ์ตามจำนวนปัจจัย พบว่า เมื่อจำนวนปัจจัยเพิ่มขึ้น อัตราความถูกต้องของเกณฑ์ทั้ง 4 เกณฑ์มีแนวโน้มลดลงอย่างต่อเนื่องทุกระดับขนาดตัวอย่าง

สรุป คือ ประสิทธิภาพในการคัดเลือกจำนวนปัจจัยสามารถแบ่งได้เป็น 2 ช่วงตามจำนวนปัจจัย ดังนี้ ช่วงแรก คือ จำนวนปัจจัยน้อย (จำนวนปัจจัยไม่เกินร้อยละ 20 ของจำนวนตัวแปร) เกณฑ์ทั้ง 4 เกณฑ์มีประสิทธิภาพในการคัดเลือกจำนวนปัจจัยสูงเท่าๆ กัน และช่วงที่สอง คือ จำนวนปัจจัยมาก (จำนวนปัจจัยมากกว่าร้อยละ 20 ของจำนวนตัวแปร) พบว่า โดยส่วนใหญ่ เกณฑ์ SIC มีประสิทธิภาพสูงที่สุด รองลงมา คือ เกณฑ์ LCV และเกณฑ์ HQ ซึ่งมีประสิทธิภาพไม่แตกต่างกัน และสุดท้าย คือ เกณฑ์ AIC ซึ่งโดยส่วนใหญ่มีประสิทธิภาพต่ำที่สุด

ตอนที่ 4.2 ผลการวิจัยของการเปรียบเทียบเกณฑ์การคัดเลือกจำนวนปัจจัยในการวิเคราะห์ปัจจัย กรณีจำนวนตัวแปรเท่ากับ 20

4.2.1 พิจารณาอัตราความถูกต้องของการคัดเลือกจำนวนปัจจัย ดังแสดงในตารางที่ 4.6-4.8 และรูปที่ 4.5-4.7

ตารางที่ 4.6 แสดงค่าอัตราความถูกต้องของเกณฑ์ LCV, เกณฑ์ AIC, เกณฑ์ SIC และเกณฑ์ HQ ในการวิเคราะห์ปัจจัย เมื่อจำนวนตัวแปรเท่ากับ 20 จำนวนปัจจัยเท่ากับ 1, 2, …, 10 และขนาดตัวอย่างเท่ากับ 300

criteria / m	LCV	AIC	SIC	HQ
1	100.00	100.00	100.00	100.00
2	100.00	100.00	100.00	100.00
3	100.00	100.00	100.00	100.00
4	100.00	100.00	100.00	100.00
5	100.00	100.00	100.00	100.00
6	99.75	98.50	100.00	100.00
7	97.50	91.00	100.00	99.50
8	92.75	82.50	100.00	92.00
9	87.50	69.25	93.25	80.00
10	72.75	56.25	81.25	66.50

หมายเหตุ ค่าอัตราความถูกต้องเป็นค่าจากการทำการวิเคราะห์ปัจจัยทั้งหมด 400 รอบ

ตารางที่ 4.7 แสดงค่าอัตราความถูกต้องของเกณฑ์ LCV, เกณฑ์ AIC, เกณฑ์ SIC และเกณฑ์ HQ ในการวิเคราะห์ปัจจัย เมื่อจำนวนตัวแปรเท่ากับ 20 จำนวนปัจจัยเท่ากับ 1, 2, ..., 10 และขนาดตัวอย่างเท่ากับ 500

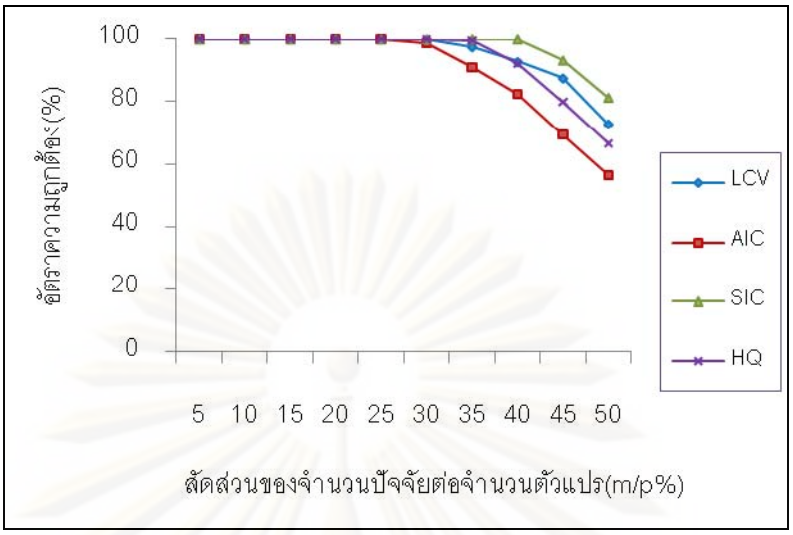
criteria / m	LCV	AIC	SIC	HQ
1	100.00	100.00	100.00	100.00
2	100.00	100.00	100.00	100.00
3	100.00	100.00	100.00	100.00
4	100.00	100.00	100.00	100.00
5	99.75	99.50	100.00	100.00
6	97.75	92.00	100.00	100.00
7	91.75	84.75	100.00	95.25
8	85.00	66.75	93.50	80.50
9	73.25	60.75	81.00	69.00
10	60.50	49.25	69.75	52.75

หมายเหตุ ค่าอัตราความถูกต้องเป็นค่าจากการทำการวิเคราะห์ปัจจัยทั้งหมด 400 รอบ

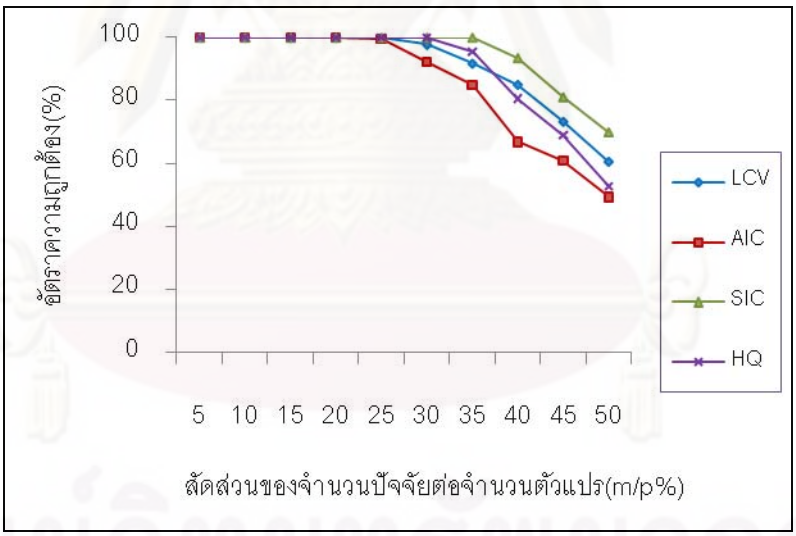
ตารางที่ 4.8 แสดงค่าอัตราความถูกต้องของเกณฑ์ LCV, เกณฑ์ AIC, เกณฑ์ SIC และเกณฑ์ HQ ในการวิเคราะห์ปัจจัย เมื่อจำนวนตัวแปรเท่ากับ 20 จำนวนปัจจัยเท่ากับ 1, 2, ..., 10 และขนาดตัวอย่างเท่ากับ 1,000

criteria / m	LCV	AIC	SIC	HQ
1	100.00	100.00	100.00	100.00
2	100.00	100.00	100.00	100.00
3	100.00	100.00	100.00	100.00
4	99.75	99.75	100.00	100.00
5	96.25	91.50	100.00	100.00
6	85.75	81.25	100.00	89.00
7	75.25	65.75	91.75	81.25
8	66.25	53.50	71.50	66.00
9	54.00	48.25	61.75	47.75
10	51.00	39.25	50.50	39.25

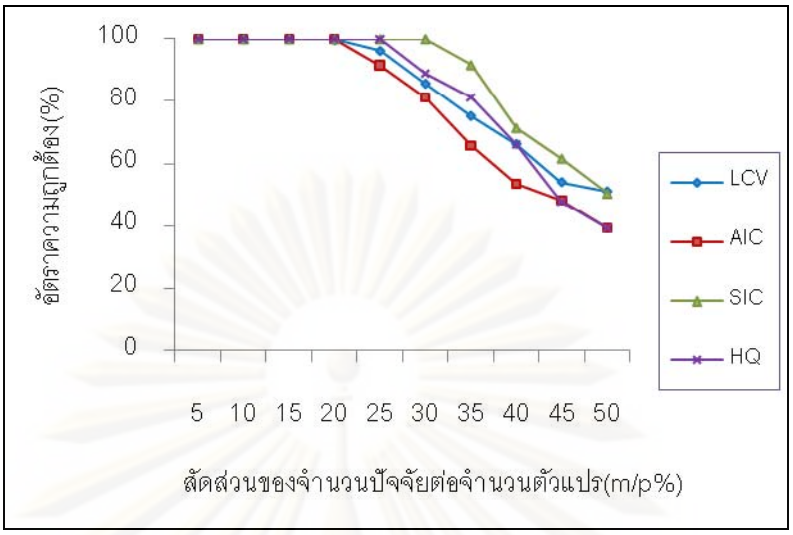
หมายเหตุ ค่าอัตราความถูกต้องเป็นค่าจากการทำการวิเคราะห์ปัจจัยทั้งหมด 400 รอบ



รูปที่ 4.5 แสดงการเปรียบเทียบเกณฑ์การคัดเลือกจำนวนปัจจัยในการวิเคราะห์ปัจจัยด้วยอัตราความถูกต้อง(%) สำหรับจำนวนตัวแปรเท่ากับ 20 และขนาดตัวอย่างเท่ากับ 300



รูปที่ 4.6 แสดงการเปรียบเทียบเกณฑ์การคัดเลือกจำนวนปัจจัยในการวิเคราะห์ปัจจัยด้วยอัตราความถูกต้อง(%) สำหรับจำนวนตัวแปรเท่ากับ 20 และขนาดตัวอย่างเท่ากับ 500



รูที่ 4.7 แสดงการเปรียบเทียบเกณฑ์การคัดเลือกจำนวนปัจจัยในการวิเคราะห์ปัจจัย ด้วยอัตราความถูกต้อง(%) สำหรับจำนวนตัวแปรเท่ากับ 20 และขนาดตัวอย่างเท่ากับ 1,000

4.2.2 **พิจารณาผลการทดสอบสมมติฐานของค่าสัดส่วนความถูกต้องของการคัดเลือกจำนวนปัจจัย ดังแสดงในตารางที่** 4.9

ตารางต่อไปนี้เป็นตารางแสดงค่า p-value สำหรับการทดสอบสมมติฐานเปรียบเทียบค่าสัดส่วนความถูกต้องของการคัดเลือกจำนวนปัจจัยในการวิเคราะห์ปัจจัย ซึ่งได้แสดงค่า p-value ไว้เพียงบางค่าเท่านั้น โดยพิจารณาจากกราฟข้างต้นแล้วพบว่า เกณฑ์การคัดเลือกจำนวนปัจจัยบางเกณฑ์มีค่าสัดส่วนความถูกต้องแตกต่างกันไม่ชัดเจน จึงได้ทำการทดสอบสมมติฐานเพื่อยืนยันผลที่ได้

ตารางที่ 4.9 แสดงค่า p-value ของการทดสอบสมมติฐานเปรียบเทียบค่าสัดส่วนของอัตราความถูกต้องของการคัดเลือกจำนวนปัจจัยในการวิเคราะห์ปัจจัย กรณีจำนวนตัวแปรเท่ากับ 20 ที่ระดับนัยสำคัญ 0.05

n	m	$H_0:p_{LCV} \leq p_{AIC}$ $H_1:p_{LCV} > p_{AIC}$	$H_0:p_{SIC} \leq p_{LCV}$ $H_1:p_{SIC} > p_{LCV}$	$H_0:p_{LCV} \leq p_{HQ}$ $H_1:p_{LCV} > p_{HQ}$	$H_0:p_{SIC} \leq p_{AIC}$ $H_1:p_{SIC} > p_{AIC}$	$H_0:p_{HQ} \leq p_{AIC}$ $H_1:p_{HQ} > p_{AIC}$	$H_0:p_{SIC} \leq p_{HQ}$ $H_1:p_{SIC} > p_{HQ}$
300	7	0.0000**	0.0007**	0.9900	0.0000**	0.0000**	0.0786
500	6	0.0001**	0.0013**	0.9987	0.0000**	0.0000**	-
1,000	8	0.0001**	0.0544	0.4702	0.0000**	0.0002**	0.0821
	10	0.0004**	0.5562	0.0004**	0.0007**	-	0.0161**

หมายเหตุ 1. ** หมายถึง การปฏิเสธสมมติฐานหลักที่ระดับนัยสำคัญ 0.05

2. ตารางนี้แสดงค่า p-value ไว้เพียงบางส่วนเท่านั้น สามารถพิจารณาค่า p-value ทั้งหมดได้ในส่วนของภาคผนวก ข

จากผลการวิจัยของการเปรียบเทียบประสิทธิภาพเกณฑ์การคัดเลือกจำนวนปัจจัยในการวิเคราะห์ปัจจัยสำหรับตัวแบบปัจจัยที่ประกอบด้วยตัวแปร 20 ตัวแปร และจำนวนปัจจัยเท่ากับ 1, 2, ..., 10 (ตารางที่ 4.6-4.9 และรูปที่ 4.5-4.7) สามารถอธิบายรายละเอียดได้ดังนี้

ขนาดตัวอย่างเท่ากับ 300

สำหรับจำนวนปัจจัยไม่เกินร้อยละ 25 ของจำนวนตัวแปร เกณฑ์การคัดเลือกจำนวนปัจจัยทั้ง 4 เกณฑ์ มีอัตราความถูกต้องสูงเท่ากัน ที่จำนวนปัจจัยมากกว่าร้อยละ 25 แต่ไม่เกินร้อยละ 35 ของจำนวนตัวแปร เกณฑ์ทั้ง 4 เกณฑ์มีอัตราความถูกต้องลดลง โดยที่เกณฑ์ SIC เกณฑ์ LCV และเกณฑ์ HQ มีอัตราความถูกต้องไม่แตกต่างกัน แต่เกณฑ์ AIC มีอัตราความถูกต้องต่ำที่สุด เมื่อจำนวนปัจจัยเพิ่มขึ้นมากกว่าร้อยละ 35 ของจำนวนตัวแปร พบว่า เกณฑ์ SIC มีอัตราความถูกต้องสูงที่สุด รองลงมา คือ เกณฑ์ LCV เกณฑ์ HQ และเกณฑ์ AIC ตามลำดับ

ขนาดตัวอย่างเท่ากับ 500

สำหรับจำนวนปัจจัยไม่เกินร้อยละ 25 ของจำนวนตัวแปร เกณฑ์การคัดเลือกจำนวนปัจจัยทั้ง 4 เกณฑ์ มีอัตราความถูกต้องสูงไม่แตกต่างกัน ที่จำนวนปัจจัยมากกว่าร้อยละ 25 แต่ไม่เกินร้อยละ 35 ของจำนวนตัวแปร เกณฑ์ทั้ง 4 เกณฑ์มีอัตราความถูกต้องลดลง โดยที่เกณฑ์ SIC มีอัตราความถูกต้องสูงที่สุด รองลงมา คือ เกณฑ์ HQ และ เกณฑ์ LCV มีอัตราความถูกต้องไม่แตกต่างกัน และเกณฑ์ AIC มีอัตราความถูกต้องต่ำที่สุด เมื่อจำนวนปัจจัยเพิ่มขึ้นมากกว่าร้อยละ 35 พบว่า เกณฑ์ SIC ยังคงมีอัตราความถูกต้องสูงที่สุด รองลงมา คือ เกณฑ์ LCV เกณฑ์ HQ และเกณฑ์ AIC ตามลำดับ

ขนาดตัวอย่างเท่ากับ 1,000

สำหรับจำนวนปัจจัยไม่เกินร้อยละ 20 ของจำนวนตัวแปร เกณฑ์การคัดเลือกจำนวนปัจจัยทั้ง 4 เกณฑ์ มีอัตราความถูกต้องสูงไม่แตกต่างกัน ที่จำนวนปัจจัยมากกว่าร้อยละ 20 แต่ไม่เกินร้อยละ 25 ของจำนวนตัวแปร เกณฑ์ SIC และเกณฑ์ HQ ยังคงมีอัตราความถูกต้องสูงที่สุด รองลงมา คือ เกณฑ์ LCV และ เกณฑ์ AIC ตามลำดับ เมื่อจำนวนปัจจัยมากกว่าร้อยละ 25 แต่ไม่เกินร้อยละ 35 ของจำนวนตัวแปร พบว่า เกณฑ์ SIC ยังคงมีอัตราความถูกต้องสูงที่สุด รองลงมา คือ เกณฑ์ HQ เกณฑ์ LCV และเกณฑ์ AIC ตามลำดับ เมื่อจำนวนปัจจัยมากกว่าร้อยละ 35 แต่ไม่เกินร้อยละ 40 ของจำนวนตัวแปร พบว่า เกณฑ์ SIC ยังคงมีอัตราความถูกต้องสูงที่สุด รองลงมา คือ เกณฑ์ LCV และเกณฑ์ HQ ซึ่งมีอัตราความถูกต้องไม่แตกต่างกัน และเกณฑ์ AIC ตามลำดับ และสำหรับจำนวนปัจจัยที่มากกว่าร้อยละ 40 ของจำนวนตัวแปร พบว่า

เกณฑ์ SIC มีอัตราความถูกต้องสูงที่สุด รองลงมา คือ เกณฑ์ LCV และเกณฑ์ AIC ซึ่งมีอัตราความถูกต้องไม่แตกต่างจากเกณฑ์ HQ

เมื่อพิจารณาอัตราความถูกต้องของเกณฑ์ทั้ง 4 เกณฑ์ตามจำนวนปัจจัย พบว่า เมื่อจำนวนปัจจัยเพิ่มขึ้น อัตราความถูกต้องของเกณฑ์ทั้ง 4 เกณฑ์มีแนวโน้มลดลงอย่างต่อเนื่องทุกระดับขนาดตัวอย่าง

สรุป คือ ประสิทธิภาพในการคัดเลือกจำนวนปัจจัยสามารถแบ่งได้เป็น 3 ช่วงโดยเฉลี่ยตามจำนวนปัจจัย ดังนี้ ช่วงที่หนึ่ง คือ จำนวนปัจจัยน้อย (จำนวนปัจจัยไม่เกินร้อยละ 23.33 ของจำนวนตัวแปร) เกณฑ์ทั้ง 4 เกณฑ์มีประสิทธิภาพในการคัดเลือกจำนวนปัจจัยสูงไม่แตกต่างกัน และช่วงที่สอง คือ จำนวนปัจจัยปานกลาง (จำนวนปัจจัยมากกว่าร้อยละ 23.33 แต่ไม่เกินร้อยละ 35 ของจำนวนตัวแปร) พบว่า โดยส่วนใหญ่เกณฑ์ SIC มีประสิทธิภาพสูงที่สุด รองลงมา คือ เกณฑ์ HQ และ เกณฑ์ LCV มีอัตราความถูกต้องไม่แตกต่างกัน และ เกณฑ์ AIC มีอัตราความถูกต้องต่ำที่สุด และช่วงที่สาม คือ จำนวนปัจจัยมาก (จำนวนปัจจัยมากกว่าร้อยละ 35 ของจำนวนตัวแปร) พบว่า โดยส่วนใหญ่เกณฑ์ SIC ยังคงมีประสิทธิภาพสูงที่สุด รองลงมา คือ เกณฑ์ LCV เกณฑ์ HQ และ เกณฑ์ AIC ตามลำดับ

ตอนที่ 4.3 ผลการวิจัยของการเปรียบเทียบเกณฑ์การคัดเลือกจำนวนปัจจัยในการวิเคราะห์ปัจจัย กรณีจำนวนตัวแปรเท่ากับ 30

4.3.1 พิจารณาอัตราความถูกต้องของการคัดเลือกจำนวนปัจจัย ดังแสดงในตารางที่ 4.10-4.11 และรูปที่ 4.8-4.9

ตารางที่ 4.10 แสดงค่าอัตราความถูกต้องของเกณฑ์ LCV, เกณฑ์ AIC, เกณฑ์ SIC และเกณฑ์ HQ ในการวิเคราะห์ปัจจัย เมื่อจำนวนตัวแปรเท่ากับ 30 จำนวนปัจจัยเท่ากับ 1, 2, ..., 15 และขนาดตัวอย่างเท่ากับ 500

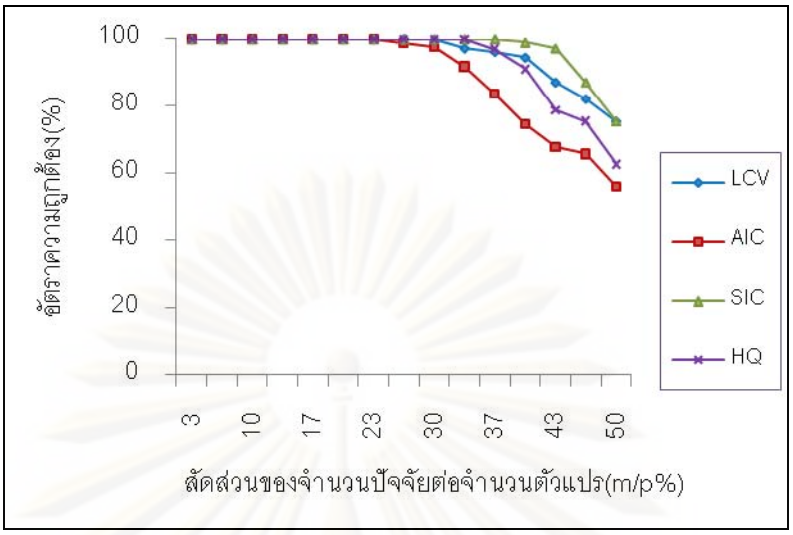
criteria / m	LCV	AIC	SIC	HQ
1	100.00	100.00	100.00	100.00
2	100.00	100.00	100.00	100.00
3	100.00	100.00	100.00	100.00
4	100.00	100.00	100.00	100.00
5	100.00	100.00	100.00	100.00
6	100.00	100.00	100.00	100.00
7	100.00	100.00	100.00	100.00
8	100.00	98.75	100.00	100.00
9	99.75	97.50	100.00	100.00
10	97.25	91.75	100.00	99.75
11	96.25	83.75	100.00	97.00
12	94.50	74.50	99.00	91.00
13	87.00	67.75	97.25	78.75
14	82.00	65.75	87.00	75.50
15	75.25	56.25	75.25	62.75

หมายเหตุ ค่าอัตราความถูกต้องเป็นค่าจากการทำการวิเคราะห์ปัจจัยทั้งหมด 400 รอบ

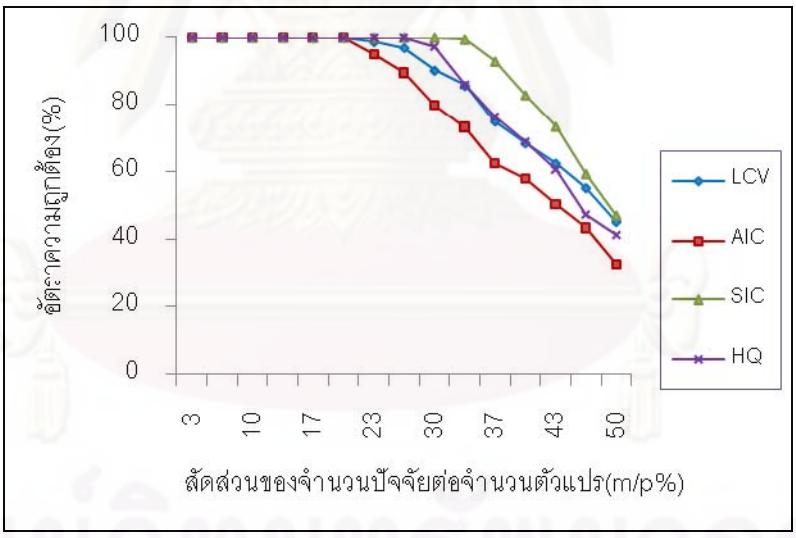
ตารางที่ 4.11 แสดงค่าอัตราความถูกต้องของเกณฑ์ LCV, เกณฑ์ AIC, เกณฑ์ SIC และเกณฑ์ HQ ในการวิเคราะห์ปัจจัย เมื่อจำนวนตัวแปรเท่ากับ 30 จำนวนปัจจัยเท่ากับ 1, 2, ..., 15 และขนาดตัวอย่างเท่ากับ 1,000

criteria / m	LCV	AIC	SIC	HQ
1	100.00	100.00	100.00	100.00
2	100.00	100.00	100.00	100.00
3	100.00	100.00	100.00	100.00
4	100.00	100.00	100.00	100.00
5	100.00	100.00	100.00	100.00
6	100.00	99.75	100.00	100.00
7	98.75	95.00	100.00	100.00
8	97.00	89.50	100.00	100.00
9	90.25	80.00	100.00	97.25
10	85.75	73.25	99.50	86.00
11	75.00	62.50	93.00	76.50
12	68.50	58.00	83.00	69.00
13	62.50	50.50	73.50	60.75
14	55.25	43.50	59.50	47.50
15	45.25	32.75	47.25	41.50

หมายเหตุ ค่าอัตราความถูกต้องเป็นค่าจากการทำการวิเคราะห์ปัจจัยทั้งหมด 400 รอบ



รูปที่ 4.8 แสดงการเปรียบเทียบเกณฑ์การคัดเลือกจำนวนปัจจัยในการวิเคราะห์ปัจจัย ด้วยอัตราความถูกต้อง(%) สำหรับจำนวนตัวแปรเท่ากับ 30 และขนาดตัวอย่างเท่ากับ 500



รูปที่ 4.9 แสดงการเปรียบเทียบเกณฑ์การคัดเลือกจำนวนปัจจัยในการวิเคราะห์ปัจจัย ด้วยอัตราความถูกต้อง(%) สำหรับจำนวนตัวแปรเท่ากับ 30 และขนาดตัวอย่างเท่ากับ 1,000

4.3.2 พิจารณาผลการทดสอบสมมติฐานของค่าสัดส่วนความถูกต้องของการคัดเลือกจำนวนปัจจัย ดังแสดงในตารางที่ 4.12

ตารางต่อไปนี้เป็นตารางแสดงค่า p-value สำหรับการทดสอบสมมติฐานเปรียบเทียบค่าสัดส่วนความถูกต้องของการคัดเลือกจำนวนปัจจัยในการวิเคราะห์ปัจจัย ซึ่งได้แสดงค่า p-value ไว้เพียงบางค่าเท่านั้น โดยพิจารณาจากกราฟข้างต้นแล้วพบว่า เกณฑ์การคัดเลือกจำนวนปัจจัยบางเกณฑ์มีค่าสัดส่วนความถูกต้องแตกต่างกันไม่ชัดเจน จึงได้ทำการทดสอบสมมติฐานเพื่อยืนยันผลที่ได้

ตารางที่ 4.12 แสดงค่า p-value ของการทดสอบสมมติฐานเปรียบเทียบค่าสัดส่วนของอัตราความถูกต้องของการคัดเลือกจำนวนปัจจัยในการวิเคราะห์ปัจจัย กรณีจำนวนตัวแปรเท่ากับ 30 ที่ระดับนัยสำคัญ 0.05

n	m	$H_0:p_{LCV} \leq p_{AIC}$ $H_1:p_{LCV} > p_{AIC}$	$H_0:p_{SIC} \leq p_{LCV}$ $H_1:p_{SIC} > p_{LCV}$	$H_0:p_{LCV} \leq p_{HQ}$ $H_1:p_{LCV} > p_{HQ}$	$H_0:p_{SIC} \leq p_{AIC}$ $H_1:p_{SIC} > p_{AIC}$	$H_0:p_{HQ} \leq p_{AIC}$ $H_1:p_{HQ} > p_{AIC}$	$H_0:p_{SIC} \leq p_{HQ}$ $H_1:p_{SIC} > p_{HQ}$
500	10	0.0003**	0.0004**	0.9982	0.0000**	0.0000**	0.1585
	11	0.0000**	0.0000**	0.7215	0.0000**	0.0000**	0.0002**
	14	0.0000**	0.0254**	0.0123**	0.0000**	0.0012**	0.0000**
	15	0.0000**	-	0.0001**	0.0000**	0.0306**	0.0001**
1,000	8	0.0000**	0.0002**	0.9998	0.0000**	0.0000**	-
	9	0.0000**	0.0000**	1.0000	0.0000**	0.0000**	0.0004**
	14	0.0004**	0.1121	0.0142**	0.0000**	0.1280	0.0003**
	15	0.0001**	0.2853	0.1423	0.0000**	0.0052**	0.0508

หมายเหตุ 1. ** หมายถึง การปฏิเสธสมมติฐานหลักที่ระดับนัยสำคัญ 0.05

2. ตารางนี้แสดงค่า p-value ไว้เพียงบางส่วนเท่านั้น สามารถพิจารณาค่า p-value ทั้งหมดได้ในส่วนของภาคผนวก ข

จากผลการวิจัยของการเปรียบเทียบประสิทธิภาพเกณฑ์การคัดเลือกจำนวนปัจจัยในการวิเคราะห์ปัจจัยสำหรับตัวแบบปัจจัยที่ประกอบด้วยตัวแปร 30 ตัวแปร และจำนวนปัจจัยเท่ากับ 1, 2, ..., 15 (ตารางที่ 4.10-4.12 และรูปที่ 4.8-4.9) สามารถอธิบายรายละเอียดได้ดังนี้

ขนาดตัวอย่างเท่ากับ 500

สำหรับจำนวนปัจจัยไม่เกินร้อยละ 23 ของจำนวนตัวแปร เกณฑ์การคัดเลือกจำนวนปัจจัยทั้ง 4 เกณฑ์ มีอัตราความถูกต้องสูงไม่แตกต่างกัน เมื่อจำนวนปัจจัยมากกว่าร้อยละ 23 แต่ไม่เกินร้อยละ 33 ของจำนวนตัวแปร เกณฑ์ทั้ง 4 เกณฑ์มีอัตราความถูกต้องลดลง โดยที่เกณฑ์ SIC เกณฑ์ HQ และ เกณฑ์ LCV มีอัตราความถูกต้องไม่แตกต่างกัน และเกณฑ์ AIC มีอัตราความถูกต้องต่ำที่สุด เมื่อจำนวนปัจจัยเพิ่มขึ้นมากกว่าร้อยละ 33 แต่ไม่เกินร้อยละ 40 ของจำนวนตัวแปร พบว่า เกณฑ์ SIC ยังคงมีอัตราความถูกต้องสูงที่สุด รองลงมา คือ เกณฑ์ LCV ซึ่งมีอัตราความถูกต้องไม่แตกต่างจากเกณฑ์ HQ และเกณฑ์ AIC ยังคงมีอัตราความถูกต้องต่ำที่สุด และเมื่อจำนวนปัจจัยมากกว่าร้อยละ 40 ของจำนวนตัวแปร พบว่า เกณฑ์ SIC ยังคงมีอัตราความถูกต้องสูงที่สุด รองลงมา คือ เกณฑ์ LCV เกณฑ์ HQ และเกณฑ์ AIC ตามลำดับ

ขนาดตัวอย่างเท่ากับ 1,000

สำหรับจำนวนปัจจัยไม่เกินร้อยละ 20 ของจำนวนตัวแปร เกณฑ์การคัดเลือกจำนวนปัจจัยทั้ง 4 เกณฑ์ มีอัตราความถูกต้องสูงไม่แตกต่างกัน ที่จำนวนปัจจัยมากกว่าร้อยละ 20 แต่ไม่เกินร้อยละ 27 ของจำนวนตัวแปร เกณฑ์ SIC และเกณฑ์ HQ ยังคงมีอัตราความถูกต้องสูงที่สุด รองลงมา คือ เกณฑ์ LCV และ เกณฑ์ AIC ตามลำดับ เมื่อจำนวนปัจจัยเพิ่มขึ้นมากกว่าร้อยละ 27 แต่ไม่เกินร้อยละ 37 ของจำนวนตัวแปร พบว่า เกณฑ์ SIC ยังคงมีอัตราความถูกต้องสูงที่สุด รองลงมา คือ เกณฑ์ HQ ซึ่งมีอัตราความถูกต้องไม่แตกต่างจากเกณฑ์ LCV และเกณฑ์ AIC มีอัตราความถูกต้องต่ำที่สุด และสำหรับจำนวนปัจจัยที่มากกว่าร้อยละ 37 พบว่า เกณฑ์ SIC และ เกณฑ์ LCV มีอัตราความถูกต้องสูงที่สุด รองลงมา คือ เกณฑ์ HQ และเกณฑ์ AIC ซึ่งมีอัตราความถูกต้องไม่แตกต่างกัน

เมื่อพิจารณาอัตราความถูกต้องของเกณฑ์ทั้ง 4 เกณฑ์ตามจำนวนปัจจัย พบว่าเมื่อจำนวนปัจจัยเพิ่มขึ้น อัตราความถูกต้องของเกณฑ์ทั้ง 4 เกณฑ์มีแนวโน้มลดลงอย่างต่อเนื่องทุกระดับขนาดตัวอย่าง

สรุป คือ ประสิทธิภาพในการคัดเลือกจำนวนปัจจัยสามารถแบ่งได้เป็น 3 ช่วงโดยเฉลี่ยตามจำนวนปัจจัย ดังนี้ ช่วงที่หนึ่ง คือ จำนวนปัจจัยน้อย (จำนวนปัจจัยไม่เกินร้อยละ 21.67 ของจำนวนตัวแปร) เกณฑ์ทั้ง 4 เกณฑ์มีประสิทธิภาพในการคัดเลือกจำนวนปัจจัยสูงไม่แตกต่าง

กัน และช่วงที่สอง คือ จำนวนปัจจัยปานกลาง (จำนวนปัจจัยมากกว่าร้อยละ 21.67 แต่ไม่เกินร้อยละ 35 ของจำนวนตัวแปร) พบว่า โดยส่วนใหญ่เกณฑ์ SIC มีประสิทธิภาพสูงที่สุด รองลงมา คือ เกณฑ์ HQ ซึ่งมีอัตราความถูกต้องไม่แตกต่างจากเกณฑ์ LCV และ เกณฑ์ AIC มีอัตราความถูกต้องต่ำที่สุด และช่วงที่สาม คือ จำนวนปัจจัยมาก (จำนวนปัจจัยมากกว่าร้อยละ 35 ของจำนวนตัวแปร) พบว่า โดยส่วนใหญ่เกณฑ์ SIC และเกณฑ์ LCV มีประสิทธิภาพสูงที่สุด รองลงมา คือ เกณฑ์ HQ และ เกณฑ์ AIC ตามลำดับ

ตอนที่ 4.4 ผลการวิจัยของการเปรียบเทียบเกณฑ์การคัดเลือกจำนวนปัจจัยในการวิเคราะห์ปัจจัย กรณีจำนวนตัวแปรเท่ากับ 40

4.4.1 พิจารณาอัตราความถูกต้องของการคัดเลือกจำนวนปัจจัย ดังแสดงในตารางที่ 4.13 และรูปที่ 4.10

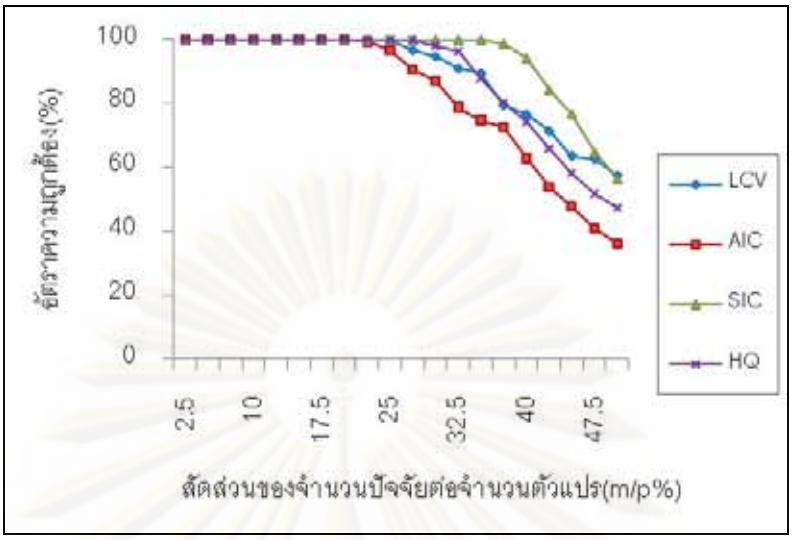
ตารางที่ 4.13 แสดงค่าอัตราความถูกต้องของเกณฑ์ LCV, เกณฑ์ AIC, เกณฑ์ SIC และเกณฑ์ HQ ในการวิเคราะห์ปัจจัย เมื่อจำนวนตัวแปรเท่ากับ 40 จำนวนปัจจัยเท่ากับ 1, 2, ..., 20 และขนาดตัวอย่างเท่ากับ 1,000

criteria / m	LCV	AIC	SIC	HQ
1	100.00	100.00	100.00	100.00
2	100.00	100.00	100.00	100.00
3	100.00	100.00	100.00	100.00
4	100.00	100.00	100.00	100.00
5	100.00	100.00	100.00	100.00
6	100.00	100.00	100.00	100.00
7	100.00	100.00	100.00	100.00
8	100.00	100.00	100.00	100.00
9	100.00	99.25	100.00	100.00
10	99.75	96.50	100.00	100.00
11	96.75	90.75	100.00	100.00
12	94.75	87.00	100.00	98.00
13	91.00	78.75	100.00	96.25
14	89.50	74.75	100.00	88.00
15	79.50	72.50	98.75	80.25

ตารางที่ 4.13(ต่อ) แสดงค่าอัตราความถูกต้องของเกณฑ์ LCV, เกณฑ์ AIC, เกณฑ์ SIC และเกณฑ์ HQ ในการวิเคราะห์ปัจจัย เมื่อจำนวนตัวแปรเท่ากับ 40 จำนวนปัจจัยเท่ากับ 1, 2, …, 20 และขนาดตัวอย่างเท่ากับ 1,000

criteria / m	LCV	AIC	SIC	HQ
16	76.50	62.75	94.25	74.25
17	71.25	53.75	84.25	65.75
18	63.50	47.75	76.75	58.00
19	62.50	40.75	64.75	51.75
20	57.25	36.00	56.25	47.25

หมายเหตุ ค่าอัตราความถูกต้องเป็นค่าจากการทำการวิเคราะห์ปัจจัยทั้งหมด 400 รอบ



รูปที่ 4.10 แสดงการเปรียบเทียบเกณฑ์การคัดเลือกจำนวนปัจจัยในการวิเคราะห์ปัจจัย ด้วยอัตราความถูกต้อง(%) สำหรับจำนวนตัวแปรเท่ากับ 40 และขนาดตัวอย่างเท่ากับ 1,000

4.4.2 พิจารณาผลการทดสอบสมมติฐานของค่าสัดส่วนความถูกต้องของการคัดเลือกจำนวนปัจจัย ดังแสดงในตารางที่ 4.12

ตารางต่อไปนี้เป็นตารางแสดงค่า p-value สำหรับการทดสอบสมมติฐานเปรียบเทียบค่าสัดส่วนความถูกต้องของการคัดเลือกจำนวนปัจจัยในการวิเคราะห์ปัจจัย ซึ่งได้แสดงค่า p-value ไว้เพียงบางค่าเท่านั้น โดยพิจารณาจากกราฟข้างต้นแล้วพบว่า เกณฑ์การคัดเลือกจำนวนปัจจัยบางเกณฑ์มีค่าสัดส่วนความถูกต้องแตกต่างกันไม่ชัดเจน จึงได้ทำการทดสอบสมมติฐานเพื่อยืนยันผลที่ได้

ตารางที่ 4.14 แสดงค่า p-value ของการทดสอบสมมติฐานเปรียบเทียบค่าสัดส่วนของอัตราความถูกต้องของการคัดเลือกจำนวนปัจจัยในการวิเคราะห์ปัจจัย กรณีจำนวนตัวแปรเท่ากับ 40 ที่ระดับนัยสำคัญ 0.05

n	m	$H_0:p_{LCV} \leq p_{AIC}$ $H_1:p_{LCV} > p_{AIC}$	$H_0:p_{SIC} \leq p_{LCV}$ $H_1:p_{SIC} > p_{LCV}$	$H_0:p_{LCV} \leq p_{HQ}$ $H_1:p_{LCV} > p_{HQ}$	$H_0:p_{SIC} \leq p_{AIC}$ $H_1:p_{SIC} > p_{AIC}$	$H_0:p_{HQ} \leq p_{AIC}$ $H_1:p_{HQ} > p_{AIC}$	$H_0:p_{SIC} \leq p_{HQ}$ $H_1:p_{SIC} > p_{HQ}$
1,000	12	0.0001**	0.0000**	0.9930	0.0000**	0.0000**	0.0022**
	13	0.0000**	0.0000**	0.9988	0.0000**	0.0000**	0.0000**
	19	0.0000**	0.2542	0.0011**	0.0000**	0.0009**	0.0001**
	20	0.0000**	0.6124	0.0023**	0.0000**	0.0006**	0.0054**

หมายเหตุ 1. ** หมายถึง การปฏิเสธสมมติฐานหลักที่ระดับนัยสำคัญ 0.05

2. ตารางนี้แสดงค่า p-value ไว้เพียงบางส่วนเท่านั้น สามารถพิจารณาค่า p-value ทั้งหมดได้ในส่วนของภาคผนวก ข

จากผลการวิจัยของการเปรียบเทียบประสิทธิภาพเกณฑ์การคัดเลือกจำนวนปัจจัยในการวิเคราะห์ปัจจัยสำหรับตัวแบบปัจจัยที่ประกอบด้วยตัวแปร 40 ตัวแปร และจำนวนปัจจัยเท่ากับ 1, 2, …, 20 (ตารางที่ 4.13-4.14 และรูปที่ 4.10) สามารถอธิบายรายละเอียดได้ดังนี้

ขนาดตัวอย่างเท่ากับ 1,000

สำหรับจำนวนปัจจัยไม่เกินร้อยละ 20 ของจำนวนตัวแปร เกณฑ์การคัดเลือกจำนวนปัจจัยทั้ง 4 เกณฑ์ มีอัตราความถูกต้องสูงไม่แตกต่างกัน ที่จำนวนปัจจัยมากกว่าร้อยละ 20 แต่ไม่เกินร้อยละ 22.5 ของจำนวนตัวแปร เกณฑ์ SIC เกณฑ์ LCV และ เกณฑ์ HQ ยังคงมีอัตราความถูกต้องสูงที่สุด แต่เกณฑ์ AIC มีอัตราความถูกต้องต่ำที่สุด เมื่อจำนวนปัจจัยเพิ่มขึ้นมากกว่าร้อยละ 22.5 แต่ไม่เกินร้อยละ 32.5 ของจำนวนตัวแปร พบว่า โดยส่วนใหญ่เกณฑ์ SIC ยังคงมีอัตราความถูกต้องสูงที่สุด รองลงมา คือ เกณฑ์ HQ เกณฑ์ LCV และเกณฑ์ AIC ตามลำดับ และสำหรับจำนวนปัจจัยมากกว่าร้อยละ 32.5 แต่ไม่เกินร้อยละ 40 ของจำนวนตัวแปร พบว่า เกณฑ์ SIC มีอัตราความถูกต้องสูงที่สุด รองลงมา คือ เกณฑ์ LCV ซึ่งมีอัตราความถูกต้องไม่แตกต่างจากเกณฑ์เกณฑ์ HQ และเกณฑ์ AIC มีอัตราความถูกต้องต่ำที่สุด และสำหรับจำนวนปัจจัยที่มากกว่าร้อยละ 40 ของจำนวนตัวแปร พบว่า โดยส่วนใหญ่เกณฑ์ SIC ยังคงมีอัตราความถูกต้องสูงที่สุด รองลงมา คือ เกณฑ์ LCV เกณฑ์ HQ และเกณฑ์ AIC ตามลำดับ

เมื่อพิจารณาอัตราความถูกต้องของเกณฑ์ทั้ง 4 เกณฑ์ตามจำนวนปัจจัย พบว่า เมื่อจำนวนปัจจัยเพิ่มขึ้น อัตราความถูกต้องของเกณฑ์ทั้ง 4 เกณฑ์มีแนวโน้มลดลงอย่างต่อเนื่องทุกระดับขนาดตัวอย่าง

สรุป คือ ประสิทธิภาพในการคัดเลือกจำนวนปัจจัยสามารถแบ่งได้เป็น 3 ช่วงตามจำนวนปัจจัย ดังนี้ ช่วงที่หนึ่ง คือ จำนวนปัจจัยน้อย (จำนวนปัจจัยไม่เกินร้อยละ 22.5 ของจำนวนตัวแปร) โดยส่วนใหญ่เกณฑ์ทั้ง 4 เกณฑ์มีประสิทธิภาพในการคัดเลือกจำนวนปัจจัยสูงไม่แตกต่างกัน และช่วงที่สอง คือ จำนวนปัจจัยปานกลาง (จำนวนปัจจัยมากกว่าร้อยละ 22.5 แต่ไม่เกินร้อยละ 37.5 ของจำนวนตัวแปร) พบว่า โดยส่วนใหญ่เกณฑ์ SIC มีประสิทธิภาพสูงที่สุด รองลงมา คือ เกณฑ์ HQ เกณฑ์ LCV และ เกณฑ์ AIC ตามลำดับ และช่วงที่สาม คือ จำนวนปัจจัยมาก (จำนวนปัจจัยมากกว่าร้อยละ 37.5 ของจำนวนตัวแปร) พบว่า โดยส่วนใหญ่เกณฑ์ SIC มีประสิทธิภาพสูงที่สุด รองลงมา คือ เกณฑ์ LCV เกณฑ์ HQ และ เกณฑ์ AIC ตามลำดับ

บทที่ 5

สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ

วัตถุประสงค์ของการวิจัยครั้งนี้ คือ การเปรียบเทียบประสิทธิภาพเกณฑ์การคัดเลือกจำนวนปัจจัยในการวิเคราะห์ปัจจัย โดยเกณฑ์ที่ใช้ในการเปรียบเทียบ ได้แก่

1. (10-fold) Likelihood Cross-Validation (LCV)
2. เกณฑ์การคัดเลือกจำนวนปัจจัยโดยใช้ข้อสนเทศของอากาอิเกะ (AIC)
3. เกณฑ์การคัดเลือกจำนวนปัจจัยโดยใช้ข้อสนเทศของชวาร์ซ (SIC)
4. เกณฑ์การคัดเลือกจำนวนปัจจัยโดยใช้ข้อสนเทศของแฮนแนนและควินน์ (HQ)

ซึ่งผู้วิจัยได้ทำการศึกษาการคัดเลือกจำนวนปัจจัยในสถานการณ์ต่างๆ ดังนี้

1. กำหนดให้จำนวนตัวแปรเท่ากับ 10 ตัวแปร จำนวนปัจจัยเท่ากับ 1,2,...,5 ปัจจัย และขนาดตัวอย่างเท่ากับ 200, 300, 500 และ 1,000
2. กำหนดให้จำนวนตัวแปรเท่ากับ 20 ตัวแปร จำนวนปัจจัยเท่ากับ 1,2,...,10 ปัจจัย และขนาดตัวอย่างเท่ากับ 300, 500 และ 1,000
3. กำหนดให้จำนวนตัวแปรเท่ากับ 30 ตัวแปร จำนวนปัจจัยเท่ากับ 1,2,...,15 ปัจจัย และขนาดตัวอย่างเท่ากับ 500 และ 1,000
4. กำหนดให้จำนวนตัวแปรเท่ากับ 40 ตัวแปร จำนวนปัจจัยเท่ากับ 1,2,...,20 ปัจจัย และขนาดตัวอย่างเท่ากับ 1,000

การเปรียบเทียบประสิทธิภาพเกณฑ์การคัดเลือกจำนวนปัจจัย ดังกล่าวข้างต้นจะพิจารณาจากอัตราความถูกต้อง (%) ของการคัดเลือกจำนวนปัจจัยได้ถูกต้องจากการทำการวิเคราะห์ปัจจัยทั้งหมด 400 รอบ และทำการทดสอบสมมติฐานเปรียบเทียบค่าสัดส่วนของอัตราความถูกต้อง เพื่อยืนยันการเปรียบเทียบประสิทธิภาพที่ได้ ซึ่งเกณฑ์การคัดเลือกจำนวนปัจจัยที่ให้ค่าอัตราความถูกต้องมากที่สุด จะเป็นเกณฑ์ที่มีประสิทธิภาพในการคัดเลือกจำนวนปัจจัยในการวิเคราะห์สูงที่สุด โดยผลการวิจัยจะเป็นการสรุปภายใต้เงื่อนไขของข้อมูลจำลอง ที่ข้อมูลจำลองมีการแจกแจงแบบปกติหลายตัวแปรที่มีเวกเตอร์ค่าเฉลี่ยเท่ากับ $\underset{\sim}{0}$ และเมทริกซ์ความแปรปรวนร่วม Σ ที่มีค่าแปรปรวนเท่ากับ 1 ซึ่งสามารถสรุปผลการวิจัยได้ดังต่อนี้

สรุปผลการวิจัย

จากการเปรียบเทียบประสิทธิภาพการคัดเลือกจำนวนปัจจัยของทั้ง 4 เกณฑ์ข้างต้น ด้วยอัตราความถูกต้องจากการคัดเลือกจำนวนปัจจัยในการวิเคราะห์ปัจจัยทั้งหมด 400 รอบและการทดสอบสมมติฐานสามารถสรุปผลได้ดังนี้

กรณีที่ 1 จำนวนตัวแปรเท่ากับ 10

จากผลการวิจัยสรุปได้ว่า การเปรียบเทียบประสิทธิภาพการคัดเลือกจำนวนปัจจัยของเกณฑ์การคัดเลือกจำนวนปัจจัยทั้ง 4 เกณฑ์ สามารถแบ่งได้เป็น 2 กรณีย่อย คือ

กรณีย่อยที่ 1 ประสิทธิภาพการคัดเลือกจำนวนปัจจัยของเกณฑ์การคัดเลือกจำนวนปัจจัยทั้ง 4 เกณฑ์ มีประสิทธิภาพสูงไม่แตกต่างกัน ซึ่งจะเกิดกรณีนี้ในช่วงที่จำนวนปัจจัยน้อย (จำนวนปัจจัยไม่เกินร้อยละ 20 ของจำนวนตัวแปร)

กรณีย่อยที่ 2 โดยส่วนใหญ่เกณฑ์การคัดเลือกจำนวนปัจจัยโดยใช้ข้อสนเทศของชวาร์ซ (SIC) มีประสิทธิภาพสูงที่สุด รองลงมา คือ เกณฑ์ 10-fold Likelihood Cross-validation (LCV) และเกณฑ์การคัดเลือกจำนวนปัจจัยโดยใช้ข้อสนเทศของแฮนแนนและควินน์ (HQ) ซึ่งมีประสิทธิภาพไม่แตกต่างกัน และสุดท้าย คือ เกณฑ์การคัดเลือกจำนวนปัจจัยโดยใช้ข้อสนเทศของอากาอิเกะ (AIC) ซึ่งโดยส่วนใหญ่มีประสิทธิภาพต่ำที่สุด ซึ่งจะเกิดกรณีนี้ในช่วงที่จำนวนปัจจัยมาก (จำนวนปัจจัยมากกว่าร้อยละ 20 ของจำนวนตัวแปร)

โดยสรุปแล้วเมื่อจำนวนปัจจัยน้อยเกณฑ์การคัดเลือกจำนวนปัจจัยแต่ละเกณฑ์มีประสิทธิภาพสูงไม่แตกต่างกัน และ เมื่อจำนวนปัจจัยเพิ่มขึ้นโดยส่วนใหญ่เกณฑ์การคัดเลือกจำนวนปัจจัยโดยใช้ข้อสนเทศของชวาร์ซ (SIC) มีประสิทธิภาพสูงที่สุด

กรณีที่ 2 จำนวนตัวแปรเท่ากับ 20, 30 และ 40

จากผลการวิจัยสรุปได้ว่า การเปรียบเทียบประสิทธิภาพการคัดเลือกจำนวนปัจจัยของเกณฑ์การคัดเลือกจำนวนปัจจัยทั้ง 4 เกณฑ์โดยเฉลี่ย สามารถแบ่งได้เป็น 3 กรณีย่อย คือ

กรณีย่อยที่ 1 ประสิทธิภาพการคัดเลือกจำนวนปัจจัยของเกณฑ์การคัดเลือกจำนวนปัจจัยทั้ง 4 เกณฑ์ มีประสิทธิภาพสูงไม่แตกต่างกัน ซึ่งจะเกิดกรณีนี้ในช่วงที่จำนวนปัจจัยน้อย (จำนวนปัจจัยไม่เกินร้อยละ 22.5 ของจำนวนตัวแปร)

กรณีย่อยที่ 2 โดยส่วนใหญ่เกณฑ์การคัดเลือกจำนวนปัจจัยโดยใช้ข้อสนเทศของชวาร์ซ (SIC) มีประสิทธิภาพสูงที่สุด รองลงมา คือ เกณฑ์การคัดเลือกจำนวนปัจจัยโดยใช้ข้อสนเทศของแฮนแนนและควินน์ (HQ) เกณฑ์ Likelihood Cross-validation (LCV) และ เกณฑ์การคัดเลือกจำนวนปัจจัยโดยใช้ข้อสนเทศของอากาอิเกะ (AIC) ตามลำดับ ซึ่งจะเกิดกรณีนี้ในช่วง

ที่จำนวนปัจจัยปานกลาง (จำนวนปัจจัยมากกว่าร้อยละ 22.5 แต่ไม่เกินร้อยละ 35.83 ของจำนวนตัวแปร)

กรณีย่อยที่ 3 โดยส่วนใหญ่เกณฑ์การคัดเลือกจำนวนปัจจัยโดยใช้ข้อสนเทศของชวาร์ช (SIC) มีประสิทธิภาพสูงที่สุด รองลงมา คือ เกณฑ์ Likelihood Cross-validation (LCV) เกณฑ์การคัดเลือกจำนวนปัจจัยโดยใช้ข้อสนเทศของแฮนแนนและควินน์ (HQ) และ เกณฑ์การคัดเลือกจำนวนปัจจัยโดยใช้ข้อสนเทศของอากาอิเกะ (AIC) ตามลำดับ ซึ่งจะเกิดกรณีนี้ในช่วงที่จำนวนปัจจัยมาก (จำนวนปัจจัยมากกว่าร้อยละ 35.83 ของจำนวนตัวแปร)

โดยสรุปแล้วเมื่อจำนวนปัจจัยน้อยเกณฑ์การคัดเลือกจำนวนปัจจัยแต่ละเกณฑ์มีประสิทธิภาพสูงไม่แตกต่างกัน และ เมื่อจำนวนปัจจัยเพิ่มขึ้นโดยส่วนใหญ่เกณฑ์การคัดเลือกจำนวนปัจจัยโดยใช้ข้อสนเทศของชวาร์ช (SIC) มีประสิทธิภาพสูงที่สุด

จากผลที่ได้ในบทที่ 4 พบว่า ผลการวิเคราะห์ในกรณีที่ 1 จำนวนตัวแปรเท่ากับ 10 ให้ผลการวิเคราะห์ใกล้เคียงกับกรณีที่ 2 จำนวนตัวแปรเท่ากับ 20, 30 และ 40 แต่เนื่องจากกรณีที่ 1 จำนวนปัจจัยที่ทำการวิเคราะห์มีน้อยจึงเห็นแนวโน้มของประสิทธิภาพการคัดเลือกจำนวนปัจจัยของทั้ง 4 เกณฑ์ ไม่ชัดเจนเท่ากรณีที่ 2 และเมื่อพิจารณาอัตราความถูกต้องตามขนาดตัวอย่างพบว่า เมื่อขนาดตัวอย่างเพิ่มขึ้น อัตราความถูกต้องของเกณฑ์การคัดเลือกจำนวนปัจจัยมีแนวโน้มลดลง ซึ่งสาเหตุเนื่องมาจากในการจำลองข้อมูลนั้น เราไม่ได้ทำการควบคุมขนาดความสัมพันธ์ระหว่างตัวแปร แต่ข้อมูลถูกจำลองมาจากทุกระดับของขนาดความสัมพันธ์ที่เป็นไปได้ทั้งหมดจากวิธี Onion ซึ่งเมทริกซ์สหสัมพันธ์ที่ได้จากวิธี Onion นั้นจะมีดีเทอร์มิแนนท์ (Determinant) เข้าใกล้ 0 (Nearly Singular Matrices) เมื่อเมทริกซ์สหสัมพันธ์มีมิติเพิ่มมากขึ้น ดังนั้นเมื่อขนาดตัวอย่างเพิ่มขึ้น อัตราความถูกต้องจึงมีแนวโน้มลดลง

อภิปรายผลการวิจัย

สำหรับผลการวิจัยเปรียบเทียบประสิทธิภาพเกณฑ์การคัดเลือกจำนวนปัจจัยในการวิเคราะห์ปัจจัยของ Michele Costa นั้นได้ผลสรุปว่า เกณฑ์ที่มีประสิทธิภาพสูงที่สุด คือ เกณฑ์ Cross-Validation (CV) รองลงมา คือ เกณฑ์การคัดเลือกจำนวนปัจจัยโดยใช้ข้อสนเทศของอากาอิเกะ (AIC) เกณฑ์การคัดเลือกจำนวนปัจจัยโดยใช้ข้อสนเทศของแฮนแนนและควินน์ (HQ) และเกณฑ์การคัดเลือกจำนวนปัจจัยโดยใช้ข้อสนเทศของชวาร์ช (SIC) ตามลำดับ ซึ่งพบว่าผลการวิจัยดังกล่าวของ Michele Costa แตกต่างจากผลการวิจัยในครั้งนี้ที่โดยส่วนใหญ่แล้วเกณฑ์การคัดเลือกจำนวนปัจจัยโดยใช้ข้อสนเทศของชวาร์ช (SIC) เป็นเกณฑ์ที่มีประสิทธิภาพสูงที่สุด รองลงมา คือ เกณฑ์การคัดเลือกจำนวนปัจจัยโดยใช้ข้อสนเทศของแฮนแนนและควินน์ (HQ) เกณฑ์ Likelihood Cross-Validation (LCV) (กรณีจำนวนปัจจัยปานกลาง) หรือ เกณฑ์

Likelihood Cross-Validation (LCV) เกณฑ์การคัดเลือกจำนวนปัจจัยโดยใช้ข้อสนเทศของแฮนแนนและควินน์ (HQ) (กรณีจำนวนปัจจัยมาก) และเกณฑ์การคัดเลือกจำนวนปัจจัยโดยใช้ข้อสนเทศของอากาอิเกะ (AIC) ตามลำดับ เนื่องจากในงานวิจัยของ Michele Costa นั้น ข้อมูลที่นำมาทำการวิเคราะห์เป็นข้อมูลที่จำลองมาจากข้อมูลจริง คือ 100 asset returns ของ Milan Stock Exchange ในปี 1986 - 1989 ซึ่งเป็นข้อมูลเพียงชุดเดียว ดังนั้นข้อจำกัดของข้อมูลสำหรับงานวิจัยของ Michele Costa คือ

1. Factor Loading ได้จากข้อมูลชุดนี้เท่านั้นซึ่งไม่ครอบคลุม Factor Loading ที่เป็นไปได้ทั้งหมด

2. ค่าเฉพาะได้จากการสุ่มให้มีการแจกแจงแบบปกติหลายตัวแปรที่มีเวกเตอร์ค่าเฉลี่ย $\underset{\sim}{0}$ และเมทริกซ์ความแปรปรวนร่วมเท่ากับเมทริกซ์เอกลักษณ์ (Identity Matrix)

ดังนั้น ผลการวิจัยของ Michele Costa จึงเหมาะสำหรับข้อมูลชุดนี้เท่านั้น แต่สำหรับงานวิจัยนี้ข้อมูลที่ได้เป็นข้อมูลจำลองที่สร้างขึ้นจากการจำลองให้ข้อมูลมีการแจกแจงแบบปกติหลายตัวแปรที่มีเวกเตอร์ค่าเฉลี่ยเท่ากับ $\underset{\sim}{0}$ และเมทริกซ์ความแปรปรวนร่วมเท่ากับ Σ ที่มีค่าความแปรปรวนเท่ากับ 1 ซึ่งเมทริกซ์ความแปรปรวนร่วมดังกล่าวได้มาจากการใช้วิธี Onion ที่เสนอโดย Ghosh และ Henderson ในปี 2003 ซึ่งเป็นวิธีในการสร้างเมทริกซ์สหสัมพันธ์ที่ได้จากการสุ่มแบบสม่ำเสมอบนเซตของเมทริกซ์สหสัมพันธ์ที่เป็นไปได้ทั้งหมด ซึ่งจะทำให้ได้

1. Factor Loading ทุกรูปแบบที่เป็นไปได้

2. ค่าเฉพาะจะแปรไปตามส่วนที่เหลือจากตัวแบบปัจจัยของตัวอย่างที่สุ่มมาซึ่งจะมีค่ามากหรือน้อยก็ได้

ดังนั้นผลการวิจัยที่ได้จึงเป็นการเฉลี่ยเหตุการณ์ที่เกิดเมทริกซ์สหสัมพันธ์ที่เป็นไปได้ทั้งหมดและมีผลแตกต่างจากผลการวิจัยของ Michele Costa

ข้อเสนอแนะ

เนื่องจากในงานวิจัยครั้งนี้ข้อมูลที่นำมาใช้ในการวิเคราะห์ถูกจำลองขึ้นจากเมทริกซ์สหสัมพันธ์ที่เป็นไปได้ทั้งหมดจากวิธี Onion ทำให้ได้เมทริกซ์สหสัมพันธ์ที่มีทุกระดับขนาดความสัมพันธ์ ผลการวิจัยที่ได้จึงเป็นการเฉลี่ยการเกิดเมทริกซ์สหสัมพันธ์ที่เป็นไปได้ทั้งหมด อย่างไรก็ตามการจำลองเมทริกซ์สหสัมพันธ์ด้วยวิธี Onion จะได้เมทริกซ์จำนวนมากที่มีดีเทอร์มิแนนท์เข้าใกล้ 0 ซึ่งเป็นคุณสมบัติของวิธี Onion ที่ได้กล่าวไว้ในงานวิจัยของ Ghosh และ Henderson [3] ดังนั้นผู้วิจัยจึงขอเสนอแนะให้ผู้ที่สนใจทำการวิจัยครั้งต่อไปศึกษาเกี่ยวกับคุณสมบัติดังกล่าวที่มีผลต่อเกณฑ์การคัดเลือกจำนวนปัจจัยเกณฑ์ต่างๆ ในงานวิจัยครั้งนี้

รายการอ้างอิง

[1] Conway, D. A., and Reinganum, M. R. January 1988. Stable Factors in Security Returns: Identification Cross-Validation. Journal of Business and Economic Statistic 6 : 1-15.

[2] Costa, M. April 1996. Factor Analysis and Information Criteria. Catalun Journal on Statistic and Operational Research 20 : 409-425.

[3] Ghosh, S., and Henderson, S G. July 2003. Behavior of the NORTA Method for Correlated Random Vector Generation as the Dimension Increases. ACM Transactions on Modeling and Computer Simulation 13 : 276-294.

[4] Ghosh, S., and Henderson, S G. January 2006. Corrigendum: Behavior of the NORTA Method for Correlated Random Vector Generation as the Dimension Increases. ACM Transactions on Modeling and Computer Simulation 16 : 93-94.

[5] กัลยา วานิชย์บัญชา. 2550. การวิเคราะห์ข้อมูลหลายตัวแปร. พิมพ์ครั้งที่2. กรุงเทพมหานคร : ธรรมสาร.

[6] Johnson, R. A., and Wichern D. W. 2002. Applied Multivariate Statistical Analysis. United States of America : Prentice-Hall.

[7] Rencher, A. C. 1995. Methods of Multivariate Analysis. New York : John Wiley.

[8] Knafl, G. J., and Grey, M. 2007. Factor Analysis Model Evaluation Through Likelihood Cross-Validation. Statistical Methods in Medical Research 16 : 77-102.

[9] Konishi, S., and Kitagawa G. 2008. Information Criteria and Statistical Modeling. Springer Series in Statistics. United States of America : Springer Science and Business Media.

ภาคผนวก

ภาคผนวก ก

โปรแกรมสำหรับงานวิจัย

โปรแกรมสำหรับเกณฑ์ 10-fold Likelihood Cross-Validation (LCV)

```
### PART 1 ONION METHOD ###

p            <-  10                  (หรือกำหนดให้ p เท่ากับ 20, 30, 40)
n            <-  200                 (หรือกำหนดให้ n เท่ากับ 300, 500, 1,000)
fullcov      <-  diag (1, p, p)
lcv          <-  c()
nfact_total  <-  matrix (, 400, 5)

for (numfact in 1:5) {

   count_LCV <-  0
   subnfact  <-  matrix (,400,1)

   for (numloop in 1:400) {

      for (k in 2:p) {

         y        <-  rbeta (1, (k-1)/2, (p-k+2)/2)
         r        <-  sqrt (y)
         z        <-  rnorm (k-1)
         normz    <-  sqrt (sum(z^2))
         theta    <-  z/normz
         w        <-  r*theta
         covk_1   <-  solve (fullcov[1:k-1, 1:k-1])
         B        <-  chol (covk_1)
         D        <-  solve (B)
         Q        <-  D%*%w
         fullcov[1:k-1,k]  <-  q
         fullcov[k,1:k-1]  <-  t (q)
         }
```

```
### PART 2 SINGULAR VALUE DECOMPOSITION ###

        SVD         <-   svd (fullcov)
        v           <-   as.matrix(SVD$u[,1:numfact])
        d           <-   diag(SVD$d[1:numfact],numfact,numfact)
        LL          <-    v%*%d%*%t (v)
        cov_noise   <-   diag (diag (fullcov) – diag (LL))
        cov_x       <-   LL + cov_noise
        mu_x        <-   rep (0,p)
        matr_x      <-   mvrnorm (n, mu_x, cov_x)
        colnames(matr_x)  <-  colnames (matr_x, do.NULL=F, prefix="X")

### PART 3 FACTOR ANALYSIS ###

        bnum        <-   10
        blength     <-   nrow (matr_x)/bnum
        matr_x.s    <-   scale (matr_x,center=T, scale=T)
        loglik.m    <-   c ()
        loglik.all  <-   c ()

        for (m in 1:(p-1)) {

            for (i in 0:(bnum-1)) {

                pred          <-   matr_x.s[-(((i*blength)+1):((i+1)*blength)),]
                pca           <-   prcomp (pred)
                sqrt_eigval   <-   diag (pca$sdev)
                loading       <-   pca$rotation %*% sqrt_eigval
                loading_m     <-   as.matrix (loading[,1:m])
                LtL           <-   loading_m %*% t (loading_m)
                mu            <-   colMeans (pred)
                err_cov       <-   diag (cov (pred)- LtL)
                specific      <-   diag (err_cov)
                covhat        <-   LtL + specific
```

```
            fit             <-   matr_x.s[((i*blength)+1):((i+1)*blength),]
            loglik.mcov <-   mvrnorm.logl (mu, covhat, fit)
            loglik.m[i+1]<-  loglik.mcov
           }
        logl.m          <-  sum (loglik.m)
        loglik.all[m]   <-  logl.m
        }
    nfact_lcv               <-  which.max (loglik.all)
    ifelse(nfact_lcv==numfact, count_LCV<-count_LCV+1, count_LCV <- count_LCV)
    subnfact[numloop,1]  <-  nfact_lcv
    }
  lcv[numfact]   <-   (count_LCV*100) / numloop
  nfact_total[, (1+(numfact-1)*1) : (numfact*1)] <-  subnfact
  }
```

โปรแกรมสำหรับเกณฑ์ Akaike's Information Criteria (AIC)

```
### PART 1 ONION METHOD ###

p           <-  10                  (หรือกำหนดให้ p เท่ากับ 20, 30, 40)
n           <-  200                 (หรือกำหนดให้ n เท่ากับ 300, 500, 1,000)
fullcov     <-  diag (1, p, p)
aic         <-  c()
nfact_total <-  matrix (, 400, 5)
for (numfact in 1:5) {
  count_AIC  <-  0
  subnfact   <-  matrix (,400,1)
  for (numloop in 1:400) {

    for (k in 2:p) {
```

```
        y       <-   rbeta (1, (k-1)/2, (p-k+2)/2)
        r       <-   sqrt (y)
        z       <-   rnorm (k-1)
        normz   <-   sqrt (sum(z^2))
        theta   <-   z/normz
        w       <-   r*theta
        covk_1  <-   solve (fullcov[1:k-1, 1:k-1])
        B       <-   chol (covk_1)
        D       <-   solve (B)
        Q       <-   D%*%w
        fullcov[1:k-1,k]  <-  q
        fullcov[k,1:k-1]  <-  t (q)
        }
### PART 2 SINGULAR VALUE DECOMPOSITION ###
      SVD         <-   svd (fullcov)
      v           <-   as.matrix(SVD$u[,1:numfact])
      d           <-   diag(SVD$d[1:numfact],numfact,numfact)
      LL          <-    v%*%d%*%t (v)
      cov_noise   <-   diag (diag (fullcov) – diag (LL))
      cov_x       <-   LL + cov_noise
      mu_x        <-   rep (0,p)
      matr_x      <-   mvrnorm (n, mu_x, cov_x)
      colnames(matr_x) <- colnames (matr_x, do.NULL=F, prefix="X")
### PART 3 FACTOR ANALYSIS ###
      matr_x.s    <-   scale(matr_x,center=T, scale=T)
      pca_forall  <-   prcomp(matr_x.s)
      sqrteigval_forall  <-  diag(pca_forall$sdev)
      loading_forall     <-  pca_forall$rotation %*% sqrteigval_forall
      mu_forall     <-   colMeans(matr_x.s)
```

```
    AIC          <-   c()

    for(m in 1:(p-1)) {

       loading_m  <-   as.matrix(loading_forall[,1:m])
       LtL        <-   loading_m %*% t(loading_m)
       errcov     <-   diag(cov(matr_x.s)- LtL)
       specific   <-   diag(errcov)
       covhat     <-   LtL + specific
       freepar    <-   p*(m+1)-(1/2)*m*(m-1)
       AIC[m]     <-   -2*mvrnorm.logl(mu_forall, covhat, matr_x.s) + 2*freepar
       }
    nfact_AIC  <-  which.min(AIC)
    ifelse(nfact_AIC==numfact, count_AIC<- count_AIC + 1, count_AIC <- count_AIC)
    subnfact[numloop,1]  <-  nfact_AIC
    }
  aic[numfact]  <-  (count_AIC*100) / numloop
  nfact_total[, (1+(numfact-1)*1) : (numfact*1)]  <- subnfact
  }
```

โปรแกรมสำหรับเกณฑ์ Schwarz's Information Criteria (SIC)

```
###PART 1 ONION METHOD ###

p          <-   10                (หรือกำหนดให้ p เท่ากับ 20, 30, 40)
n          <-   200               (หรือกำหนดให้ n เท่ากับ 300, 500, 1,000)
fullcov    <-   diag (1, p, p)
sic        <-   c()
nfact_total  <-  matrix (, 400, 5)

for (numfact in 1:5) {

   count_SIC  <-   0
```

```
    subnfact    <-   matrix (,400,1)

    for (numloop in 1:400) {

        for (k in 2:p) {

            y          <-   rbeta (1, (k-1)/2, (p-k+2)/2)
            r          <-   sqrt (y)
            z          <-   rnorm (k-1)
            normz      <-   sqrt (sum(z^2))
            theta      <-   z/normz
            w          <-   r*theta
            covk_1     <-   solve (fullcov[1:k-1, 1:k-1])
            B          <-   chol (covk_1)
            D          <-   solve (B)
            Q          <-   D%*%w
            fullcov[1:k-1,k]  <-  q
            fullcov[k,1:k-1]  <-  t (q)
            }

### PART 2 SINGULAR VALUE DECOMPOSITION ###

        SVD          <-   svd (fullcov)
        v            <-   as.matrix(SVD$u[,1:numfact])
        d            <-   diag(SVD$d[1:numfact],numfact,numfact)
        LL           <-    v%*%d%*%t (v)
        cov_noise    <-   diag (diag (fullcov) – diag (LL))
        cov_x        <-   LL + cov_noise
        mu_x         <-   rep (0,p)
        matr_x       <-   mvrnorm (n, mu_x, cov_x)
        colnames(matr_x) <- colnames (matr_x, do.NULL=F, prefix="X")

### PART 3 FACTOR ANALYSIS ###

        matr_x.s          <-    scale(matr_x,center=T, scale=T)
```

```
    pca_forall      <-  prcomp(matr_x.s)   sqrteigval_forall <- diag(pca_forall$sdev)
    loading_forall  <-  pca_forall$rotation %*% sqrteigval_forall
    mu_forall       <-  colMeans(matr_x.s)
    SIC             <-  c()
    for(m in 1:(p-1)) {
      loading_m  <-  as.matrix(loading_forall[,1:m])
      LtL        <-  loading_m %*% t(loading_m)
      errcov     <-  diag(cov(matr_x.s)- LtL)
      specific   <-  diag(errcov)
      covhat     <-  LtL + specific
      freepar    <-  p*(m+1)-(1/2)*m*(m-1)
      SIC[m]     <-  -2*mvrnorm.logl(mu_forall,covhat,matr_x.s)+freepar*log(n,exp(1))
      }
    nfact_SIC  <-  which.min(SIC)
    ifelse(nfact_SIC==numfact, count_SIC <- count_SIC + 1, count_SIC <- count_SIC)
    subnfact[numloop,1]  <-  nfact_SIC
    }
  sic[numfact]  <-  (count_SIC*100) / numloop
  nfact_total[, (1+(numfact-1)*1) : (numfact*1)]  <- subnfact
  }
```

โปรแกรมสำหรับเกณฑ์ Hannan and Quinn's Information Criteria (HQ)

```
###PART 1 ONION METHOD ###
p           <-  10                  (หรือกำหนดให้ p เท่ากับ 20, 30, 40)
fullcov     <-  diag (1, p, p)
n           <-  200                 (หรือกำหนดให้ n เท่ากับ 300, 500, 1,000)
hq          <-  c()
nfact_total <-  matrix (, 400, 5)

for (numfact in 1:5) {
```

```
count_HQ   <-   0
subnfact    <-   matrix (,400,1)
for (numloop in 1:400) {
   for (k in 2:p) {
      y          <-   rbeta (1, (k-1)/2, (p-k+2)/2)
      r          <-   sqrt (y)
      z          <-   rnorm (k-1)
      normz      <-   sqrt (sum(z^2))
      theta      <-   z/normz
      w          <-   r*theta
      covk_1     <-   solve (fullcov[1:k-1, 1:k-1])
      B          <-   chol (covk_1)
      D          <-   solve (B)
      Q          <-   D%*%w
      fullcov[1:k-1,k]  <-  q
      fullcov[k,1:k-1]  <-  t (q)
      }
### PART 2 SINGULAR VALUE DECOMPOSITION ###
      SVD        <-   svd (fullcov)
      v          <-   as.matrix(SVD$u[,1:numfact])
      d          <-   diag(SVD$d[1:numfact],numfact,numfact)
      LL         <-    v%*%d%*%t (v)
      cov_noise  <-   diag (diag (fullcov) – diag (LL))
      cov_x      <-   LL + cov_noise
      mu_x       <-   rep (0,p)
      matr_x     <-   mvrnorm (n, mu_x, cov_x)
      colnames(matr_x) <- colnames (matr_x, do.NULL=F, prefix="X")
```

```
### PART 3 FACTOR ANALYSIS ###

      matr_x.s        <-   scale(matr_x,center=T, scale=T)
      pca_forall      <-   prcomp(matr_x.s)   sqrteigval_forall  <-  diag(pca_forall$sdev)
      loading_forall  <-   pca_forall$rotation %*% sqrteigval_forall
      mu_forall       <-   colMeans(matr_x.s)
      HQ              <-   c()
      for(m in 1:(p-1)) {
         loading_m    <-   as.matrix(loading_forall[,1:m])
         LtL          <-   loading_m %*% t(loading_m)
         errcov       <-   diag(cov(matr_x.s)- LtL)
         specific     <-   diag(errcov)
         covhat       <-   LtL + specific
         freepar      <-   p*(m+1)-(1/2)*m*(m-1)
         HQ[m]        <-   -2*mvrnorm.logl(mu_forall,covhat,matr_x.s)+
                           2*freepar*log(log(n,exp(1)),exp(1))
         }
      nfact_HQ  <-  which.min(HQ)
      ifelse(nfact_HQ==numfact, count_HQ <- count_HQ + 1, count_HQ <- count_HQ)
      subnfact[numloop,1]  <-  nfact_HQ
      }
   hq[numfact]  <-  (count_HQ*100) / numloop
   nfact_total[, (1+(numfact-1)*1) : (numfact*1)]  <- subnfact
   }
```

โปรแกรมของฟังก์ชัน mvrnorm.logl

```
mvrnorm.logl  <-  function(mu,covhat,x){
n             <-   nrow(x)
p             <-   ncol(x)
meanx         <-   matrix(rep(mu,each=n),n,p)
```

```
centerx     <-  as.matrix(x-meanx)
lterm       <-  centerx %*% solve(covhat) %*% t(centerx)
sum_lterm   <-  sum(diag(lterm))
logl        <-  (-n*p/2)*log(2*pi,base=exp(1))-(n/2)*log(det(covhat),base=exp(1))-
                0.5*sum_lterm
return(logl)
}
```

ภาคผนวก ข

ตารางที่ 1 แสดงค่า p-value ของการทดสอบสมมติฐานเปรียบเทียบค่าสัดส่วนความถูกต้องของการคัดเลือกจำนวนปัจจัยในการวิเคราะห์ปัจจัย กรณีจำนวนตัวแปรเท่ากับ 10 ขนาดตัวอย่างเท่ากับ 200 ที่ระดับนัยสำคัญ 0.05

m	$H_0:P_{LCV} \leq P_{AIC}$ $H_1:P_{LCV} > P_{AIC}$	$H_0:P_{SIC} \leq P_{LCV}$ $H_1:P_{SIC} > P_{LCV}$	$H_0:P_{LCV} \leq P_{HQ}$ $H_1:P_{LCV} > P_{HQ}$	$H_0:P_{SIC} \leq P_{AIC}$ $H_1:P_{SIC} > P_{AIC}$	$H_0:P_{HQ} \leq P_{AIC}$ $H_1:P_{HQ} > P_{AIC}$	$H_0:P_{SIC} \leq P_{HQ}$ $H_1:P_{SIC} > P_{HQ}$
1	-	-	-	-	-	-
2	-	-	-	-	-	-
3	0.0413**	-	-	0.0413**	0.0413**	-
4	0.0000**	0.0006**	0.1377	0.0000**	0.0000**	0.0000**
5	0.0034**	0.0077**	0.2503	0.0000**	0.0208**	0.0010**

หมายเหตุ – หมายถึง ค่าสัดส่วนความถูกต้องของเกณฑ์ที่ทำการเปรียบเทียบเท่ากัน

ตารางที่ 2 แสดงค่า p-value ของการทดสอบสมมติฐานเปรียบเทียบค่าสัดส่วนความถูกต้องของการคัดเลือกจำนวนปัจจัยในการวิเคราะห์ปัจจัย กรณีจำนวนตัวแปรเท่ากับ 10 ขนาดตัวอย่างเท่ากับ 300 ที่ระดับนัยสำคัญ 0.05

m	$H_0:P_{LCV} \leq P_{AIC}$ $H_1:P_{LCV} > P_{AIC}$	$H_0:P_{SIC} \leq P_{LCV}$ $H_1:P_{SIC} > P_{LCV}$	$H_0:P_{LCV} \leq P_{HQ}$ $H_1:P_{LCV} > P_{HQ}$	$H_0:P_{SIC} \leq P_{AIC}$ $H_1:P_{SIC} > P_{AIC}$	$H_0:P_{HQ} \leq P_{AIC}$ $H_1:P_{HQ} > P_{AIC}$	$H_0:P_{SIC} \leq P_{HQ}$ $H_1:P_{SIC} > P_{HQ}$
1	-	-	-	-	-	-
2	-	-	-	-	-	-
3	0.1223	0.0225**	0.9108	0.0022**	0.0095**	0.1585
4	0.0000**	0.0005**	0.0084**	0.0000**	0.0024**	0.0000**
5	0.0003**	0.2440	0.0040**	0.0000**	0.2151	0.0004**

หมายเหตุ – หมายถึง ค่าสัดส่วนความถูกต้องของเกณฑ์ที่ทำการเปรียบเทียบเท่ากัน

ตารางที่ 3 แสดงค่า p-value ของการทดสอบสมมติฐานเปรียบเทียบค่าสัดส่วนความถูกต้องของการคัดเลือกจำนวนปัจจัยในการวิเคราะห์ปัจจัย กรณีจำนวนตัวแปรเท่ากับ 10 ขนาดตัวอย่างเท่ากับ 500 ที่ระดับนัยสำคัญ 0.05

m	$H_0:P_{LCV} \leq P_{AIC}$ $H_1:P_{LCV} > P_{AIC}$	$H_0:P_{SIC} \leq P_{LCV}$ $H_1:P_{SIC} > P_{LCV}$	$H_0:P_{LCV} \leq P_{HQ}$ $H_1:P_{LCV} > P_{HQ}$	$H_0:P_{SIC} \leq P_{AIC}$ $H_1:P_{SIC} > P_{AIC}$	$H_0:P_{HQ} \leq P_{AIC}$ $H_1:P_{HQ} > P_{AIC}$	$H_0:P_{SIC} \leq P_{HQ}$ $H_1:P_{SIC} > P_{HQ}$
1	-	-	-	-	-	-
2	0.1585	-	-	0.1585	0.1585	-
3	0.0557	0.0000**	0.9974	0.0000**	0.0000**	0.0288**
4	0.0094**	0.0454**	0.4325	0.0000**	0.0146**	0.0314**
5	0.0448**	0.4717	0.6651	0.0385**	0.0169**	0.6389

หมายเหตุ – หมายถึง ค่าสัดส่วนความถูกต้องของเกณฑ์ที่ทำการเปรียบเทียบเท่ากัน

ตารางที่ 4 แสดงค่า p-value ของการทดสอบสมมติฐานเปรียบเทียบค่าสัดส่วนความถูกต้องของการคัดเลือกจำนวนปัจจัยในการวิเคราะห์ปัจจัย กรณีจำนวนตัวแปรเท่ากับ 10 ขนาดตัวอย่างเท่ากับ 1,000 ที่ระดับนัยสำคัญ 0.05

m	$H_0:P_{LCV} \leq P_{AIC}$ $H_1:P_{LCV} > P_{AIC}$	$H_0:P_{SIC} \leq P_{LCV}$ $H_1:P_{SIC} > P_{LCV}$	$H_0:P_{LCV} \leq P_{HQ}$ $H_1:P_{LCV} > P_{HQ}$	$H_0:P_{SIC} \leq P_{AIC}$ $H_1:P_{SIC} > P_{AIC}$	$H_0:P_{HQ} \leq P_{AIC}$ $H_1:P_{HQ} > P_{AIC}$	$H_0:P_{SIC} \leq P_{HQ}$ $H_1:P_{SIC} > P_{HQ}$
1	-	-	-	-	-	-
2	-	-	-	-	-	-
3	0.0426**	0.0000**	0.8209	0.0000**	0.0044**	0.0001**
4	0.2730	0.0056**	0.2984	0.0009**	0.4701	0.0011**
5	0.6897	0.0785	0.5000	0.1787	0.6897	0.0785

หมายเหตุ – หมายถึง ค่าสัดส่วนความถูกต้องของเกณฑ์ที่ทำการเปรียบเทียบเท่ากัน

ตารางที่ 5 แสดงค่า p-value ของการทดสอบสมมติฐานเปรียบเทียบค่าสัดส่วนความถูกต้องของการคัดเลือกจำนวนปัจจัยในการวิเคราะห์ปัจจัย กรณีจำนวนตัวแปรเท่ากับ 20 ขนาดตัวอย่างเท่ากับ 300 ที่ระดับนัยสำคัญ 0.05

m	$H_0:P_{LCV} \leq P_{AIC}$ $H_1:P_{LCV} > P_{AIC}$	$H_0:P_{SIC} \leq P_{LCV}$ $H_1:P_{SIC} > P_{LCV}$	$H_0:P_{LCV} \leq P_{HQ}$ $H_1:P_{LCV} > P_{HQ}$	$H_0:P_{SIC} \leq P_{AIC}$ $H_1:P_{SIC} > P_{AIC}$	$H_0:P_{HQ} \leq P_{AIC}$ $H_1:P_{HQ} > P_{AIC}$	$H_0:P_{SIC} \leq P_{HQ}$ $H_1:P_{SIC} > P_{HQ}$
1	-	-	-	-	-	-
2	-	-	-	-	-	-
3	-	-	-	-	-	-
4	-	-	-	-	-	-
5	-	-	-	-	-	-
6	0.0288**	0.1585	0.8415	0.0070**	0.0070**	-
7	0.0000**	0.0007**	0.9900	0.0000**	0.0000**	0.0786
8	0.0000**	0.0000**	0.3447	0.0000**	0.0000**	0.0000**
9	0.0000**	0.0029**	0.0020**	0.0000**	0.0002**	0.0000**
10	0.0000**	0.0021**	0.0273**	0.0000**	0.0015**	0.0000**

หมายเหตุ – หมายถึง ค่าสัดส่วนความถูกต้องของเกณฑ์ที่ทำการเปรียบเทียบเท่ากัน

ตารางที่ 6 แสดงค่า p-value ของการทดสอบสมมติฐานเปรียบเทียบค่าสัดส่วนความถูกต้องของการคัดเลือกจำนวนปัจจัยในการวิเคราะห์ปัจจัย กรณีจำนวนตัวแปรเท่ากับ 20 ขนาดตัวอย่างเท่ากับ 500 ที่ระดับนัยสำคัญ 0.05

m	$H_0:P_{LCV} \leq P_{AIC}$ $H_1:P_{LCV} > P_{AIC}$	$H_0:P_{SIC} \leq P_{LCV}$ $H_1:P_{SIC} > P_{LCV}$	$H_0:P_{LCV} \leq P_{HQ}$ $H_1:P_{LCV} > P_{HQ}$	$H_0:P_{SIC} \leq P_{AIC}$ $H_1:P_{SIC} > P_{AIC}$	$H_0:P_{HQ} \leq P_{AIC}$ $H_1:P_{HQ} > P_{AIC}$	$H_0:P_{SIC} \leq P_{HQ}$ $H_1:P_{SIC} > P_{HQ}$
1	-	-	-	-	-	-
2	-	-	-	-	-	-
3	-	-	-	-	-	-
4	-	-	-	-	-	-
5	0.2815	0.1585	0.8415	0.0784	0.0784	-
6	0.0001**	0.0013**	0.9987	0.0000**	0.0000**	-
7	0.0011**	0.0000**	0.9777	0.0000**	0.0000**	0.0000**
8	0.0000**	0.0001**	0.0461**	0.0000**	0.0000**	0.0000**
9	0.0001**	0.0045**	0.0924	0.0000**	0.0073**	0.0003**
10	0.0007**	0.0030**	0.0135**	0.0000**	0.1611	0.0001**

หมายเหตุ – หมายถึง ค่าสัดส่วนความถูกต้องของเกณฑ์ที่ทำการเปรียบเทียบเท่ากัน

ตารางที่ 7 แสดงค่า p-value ของการทดสอบสมมติฐานเปรียบเทียบค่าสัดส่วนความถูกต้องของการคัดเลือกจำนวนปัจจัยในการวิเคราะห์ปัจจัย กรณีจำนวนตัวแปรเท่ากับ 20 ขนาดตัวอย่างเท่ากับ 1,000 ที่ระดับนัยสำคัญ 0.05

m	$H_0:P_{LCV} \leq P_{AIC}$ $H_1:P_{LCV} > P_{AIC}$	$H_0:P_{SIC} \leq P_{LCV}$ $H_1:P_{SIC} > P_{LCV}$	$H_0:P_{LCV} \leq P_{HQ}$ $H_1:P_{LCV} > P_{HQ}$	$H_0:P_{SIC} \leq P_{AIC}$ $H_1:P_{SIC} > P_{AIC}$	$H_0:P_{HQ} \leq P_{AIC}$ $H_1:P_{HQ} > P_{AIC}$	$H_0:P_{SIC} \leq P_{HQ}$ $H_1:P_{SIC} > P_{HQ}$
1	-	-	-	-	-	-
2	-	-	-	-	-	-
3	-	-	-	-	-	-
4	-	0.1585	0.8415	0.1585	0.1585	-
5	0.0025**	0.0000**	1.0000	0.0000**	0.0000**	-
6	0.0432**	0.0000**	0.9168	0.0000**	0.0010**	0.0000**
7	0.0016**	0.0000**	0.9801	0.0000**	0.0000**	0.0000**
8	0.0001**	0.0544	0.4702	0.0000**	0.0002**	0.0821
9	0.0519	0.0132**	0.0385**	0.0001**	0.5563	0.0016**
10	0.0004**	0.5562	0.0004**	0.0007**	-	0.0161**

หมายเหตุ – หมายถึง ค่าสัดส่วนความถูกต้องของเกณฑ์ที่ทำการเปรียบเทียบเท่ากัน

ตารางที่ 8 แสดงค่า p-value ของการทดสอบสมมติฐานเปรียบเทียบค่าสัดส่วนความถูกต้องของการคัดเลือกจำนวนปัจจัยในการวิเคราะห์ปัจจัย กรณีจำนวนตัวแปรเท่ากับ 30 ขนาดตัวอย่างเท่ากับ 500 ที่ระดับนัยสำคัญ 0.05

m	$H_0:P_{LCV} \leq P_{AIC}$ $H_1:P_{LCV} > P_{AIC}$	$H_0:P_{SIC} \leq P_{LCV}$ $H_1:P_{SIC} > P_{LCV}$	$H_0:P_{LCV} \leq P_{HQ}$ $H_1:P_{LCV} > P_{HQ}$	$H_0:P_{SIC} \leq P_{AIC}$ $H_1:P_{SIC} > P_{AIC}$	$H_0:P_{HQ} \leq P_{AIC}$ $H_1:P_{HQ} > P_{AIC}$	$H_0:P_{SIC} \leq P_{HQ}$ $H_1:P_{SIC} > P_{HQ}$
1	-	-	-	-	-	-
2	-	-	-	-	-	-
3	-	-	-	-	-	-
4	-	-	-	-	-	-
5	-	-	-	-	-	-
6	-	-	-	-	-	-
7	-	-	-	-	-	-
8	0.0124**	-	-	0.0124**	0.0124**	-
9	0.0031**	0.1585	0.8415	0.0007**	0.0007**	-
10	0.0003**	0.0004**	0.9982	0.0000**	0.0000**	0.1585
11	0.0000**	0.0000**	0.7215	0.0000**	0.0000**	0.0002**
12	0.0000**	0.0002**	0.0281	0.0000**	0.0000**	0.0000**
13	0.0000**	0.0000**	0.0010**	0.0000**	0.0002**	0.0000**
14	0.0000**	0.0254**	0.0123**	0.0000**	0.0012**	0.0000**
15	0.0000**	-	0.0001**	0.0000**	0.0306**	0.0001**

หมายเหตุ – หมายถึง ค่าสัดส่วนความถูกต้องของเกณฑ์ที่ทำการเปรียบเทียบเท่ากัน

ตารางที่ 9 แสดงค่า p-value ของการทดสอบสมมติฐานเปรียบเทียบค่าสัดส่วนความถูกต้องของการคัดเลือกจำนวนปัจจัยในการวิเคราะห์ปัจจัย กรณีจำนวนตัวแปรเท่ากับ 30 ขนาดตัวอย่างเท่ากับ 1,000 ที่ระดับนัยสำคัญ 0.05

m	$H_0:P_{LCV} \leq P_{AIC}$ $H_1:P_{LCV} > P_{AIC}$	$H_0:P_{SIC} \leq P_{LCV}$ $H_1:P_{SIC} > P_{LCV}$	$H_0:P_{LCV} \leq P_{HQ}$ $H_1:P_{LCV} > P_{HQ}$	$H_0:P_{SIC} \leq P_{AIC}$ $H_1:P_{SIC} > P_{AIC}$	$H_0:P_{HQ} \leq P_{AIC}$ $H_1:P_{HQ} > P_{AIC}$	$H_0:P_{SIC} \leq P_{HQ}$ $H_1:P_{SIC} > P_{HQ}$
1	-	-	-	-	-	-
2	-	-	-	-	-	-
3	-	-	-	-	-	-
4	-	-	-	-	-	-
5	-	-	-	-	-	-
6	0.1585	-	-	0.1585	0.1585	-
7	0.0012**	0.0124**	0.9876	0.0000**	0.0000**	-
8	0.0000**	0.0002**	0.9998	0.0000**	0.0000**	-
9	0.0000**	0.0000**	1.0000	0.0000**	0.0000**	0.0004**
10	0.0000**	0.0000**	0.5404	0.0000**	0.0000**	0.0000**
11	0.0001**	0.0000**	0.6897	0.0000**	0.0000**	0.0000**
12	0.0010**	0.0000**	0.5606	0.0000**	0.0006**	0.0000**
13	0.0003**	0.0004**	0.3054	0.0000**	0.0018**	0.0001**
14	0.0004**	0.1121	0.0142**	0.0000**	0.1280	0.0003**
15	0.0001**	0.2853	0.1423	0.0000**	0.0052**	0.0508

หมายเหตุ – หมายถึง ค่าสัดส่วนความถูกต้องของเกณฑ์ที่ทำการเปรียบเทียบเท่ากัน

ตารางที่ 10 แสดงค่า p-value ของการทดสอบสมมติฐานเปรียบเทียบค่าสัดส่วนความถูกต้องของการคัดเลือกจำนวนปัจจัยในการวิเคราะห์ปัจจัย กรณีจำนวนตัวแปรเท่ากับ 40 ขนาดตัวอย่างเท่ากับ 1,000 ที่ระดับนัยสำคัญ 0.05

m	$H_0:P_{LCV} \leq P_{AIC}$ $H_1:P_{LCV} > P_{AIC}$	$H_0:P_{SIC} \leq P_{LCV}$ $H_1:P_{SIC} > P_{LCV}$	$H_0:P_{LCV} \leq P_{HQ}$ $H_1:P_{LCV} > P_{HQ}$	$H_0:P_{SIC} \leq P_{AIC}$ $H_1:P_{SIC} > P_{AIC}$	$H_0:P_{HQ} \leq P_{AIC}$ $H_1:P_{HQ} > P_{AIC}$	$H_0:P_{SIC} \leq P_{HQ}$ $H_1:P_{SIC} > P_{HQ}$
1	-	-	-	-	-	-
2	-	-	-	-	-	-
3	-	-	-	-	-	-
4	-	-	-	-	-	-
5	-	-	-	-	-	-
6	-	-	-	-	-	-
7	-	-	-	-	-	-
8	-	-	-	-	-	-
9	0.0413**	-	-	0.0413**	0.0413**	-
10	0.0004**	0.1585	0.8415	0.0001**	0.0001**	-
11	0.0002**	0.0001**	0.9999	0.0000**	0.0000**	-
12	0.0001**	0.0000**	0.9930	0.0000**	0.0000**	0.0022**
13	0.0000**	0.0000**	0.9988	0.0000**	0.0000**	0.0000**
14	0.0000**	0.0000**	0.2510	0.0000**	0.0000**	0.0000**
15	0.0102**	0.0000**	0.6043	0.0000**	0.0049**	0.0000**
16	0.0000**	0.0000**	0.2301	0.0000**	0.0002**	0.0000**
17	0.0000**	0.0000**	0.0470**	0.0000**	0.0003**	0.0000**
18	0.0000**	0.0000**	0.0556	0.0000**	0.0018**	0.0000**
19	0.0000**	0.2542	0.0011**	0.0000**	0.0009**	0.0001**
20	0.0000**	0.6124	0.0023**	0.0000**	0.0006**	0.0054**

หมายเหตุ – หมายถึง ค่าสัดส่วนความถูกต้องของเกณฑ์ที่ทำการเปรียบเทียบเท่ากัน

ประวัติผู้เขียนวิทยานิพนธ์

นางสาวพรรณนิภา รินทระ เกิดเมื่อวันที่ 12 มกราคม 2527 สำเร็จการศึกษาระดับมัธยมศึกษาจากโรงเรียนปิยะมหาราชาลัย จังหวัดนครพนม เข้าศึกษาในคณะวิทยาศาสตร์ มหาวิทยาลัยธรรมศาสตร์ และสำเร็จปริญญาวิทยาศาสตรบัณฑิต (คณิตศาสตร์) และเข้าศึกษาหลักสูตรสถิติศาสตรมหาบัณฑิต จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2550