บทที่ 1

บทนำ

ความเป็นมาและความสำคัญของปัญหา

เทคนิคทางสถิติที่ใช้ในการวิเคราะห์ข้อมูลเมื่อตัวแปรเป็นตัวแปรพหุ(Multivariate)ที่ได้รับการกล่าวขานกันอยู่เสมอ คือ วิธีการวิเคราะห์ตัวประกอบ(Factor Analysis Methods) ซึ่งก็คือวิธีการค้นหาชุดของตัวแปรที่สามารถอธิบายมิติ (Dimensions) หรือ ตัวประกอบ (Factors) จำนวนหนึ่งที่น้อยกว่าจำนวนตัวแปรพร้อมทั้งค้นหาว่าอะไรคือตัวประกอบนั้น การวิเคราะห์ตัวประกอบนี้เป็นเทคนิคทางสถิติที่มีวัตถุประสงค์ในการลดปริมาณของตัวแปรให้มีจำนวนน้อยลง (Variable reduction)โดยอาศัยโครงสร้างและแบบแผนของความสัมพันธ์ที่มีระหว่างตัวแปร ทั้งนี้เพื่อให้ง่ายต่อความเข้าใจ และทำให้ทราบถึงโครงสร้างและแบบแผนของข้อมูล (Structure and pattern of data)นั้น การวิเคราะห์ตัวประกอบจะถูกนำมาใช้ในค้นหาว่า ความสัมพันธ์ของตัวแปรนั้นจะสามารถอธิบายด้วยตัวแปรตามสมมติฐานหรือตัวประกอบที่มีจำนวนน้อยลงได้หรือไม่

โดยยึดหลักที่ว่าการที่ตัวแปรมีความสัมพันธ์กันนั้นก็เนื่องจากว่าตัวแปรต่าง ๆ เหล่านี้มีตัวประกอบร่วมกัน (Common factors) สามารถสังเกตได้จากการจับกลุ่มกันของตัวแปรหรือค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรนั้นมีความสัมพันธ์กันในระดับสูง

ในการวิจัยทางสังคมศาสตร์นั้น การใช้เทคนิคการวิเคราะห์ตัวประกอบนี้มีจุดมุ่งหมายที่สำคัญอยู่ 3 ประการ คือ (สุชาติ ประสิทธิ์รัฐสินธุ์, 2527)

1. เพื่อแสวงหาตัวประกอบร่วมที่อธิบายถึงความสัมพันธ์ร่วมกันระหว่างตัวแปรต่างๆ โดยที่จำนวนของตัวประกอบร่วมที่หาได้นั้นจะต้องมีจำนวนน้อยกว่าจำนวนตัวแปรซึ่งจะถูกเรียกว่า Exploratory factor analysis

2. เพื่อพิสูจน์ สนับสนุน ตรวจสอบสมมติฐานเกี่ยวกับโครงสร้างของข้อมูล หรือตัวแปรว่ามีตัวประกอบร่วมกันกี่ตัวประกอบ อะไรบ้าง และตัวแปรแต่ละตัวมีความสัมพันธ์กับตัวประกอบอย่างไร ซึ่งเรียกว่า Confirmatory factor analysis

3. เพื่อนำเอาโครงสร้างของความสัมพันธ์ระหว่างตัวประกอบ และตัวแปรเหล่านี้ไปใช้สร้างคะแนนองค์ประกอบ (Factor score) คะแนนที่ได้นี้เปรียบเสมือนค่าของตัวแปรตัวใหม่ที่ประกอบด้วยตัวแปรเดิมหลาย ๆ ตัวนี้ จะถูกเรียกว่า Composite variable

ระเบียบวิธีการวิเคราะห์ตัวประกอบนั้นลินเดอร์แมนและคณะ(Lindeman , Merenda, and Gold , 1980) ได้สรุปไว้ว่าลักษณะของการวิเคราะห์ตัวประกอบแบบ Exploratory (Exploratary factor analysis)นี้ จะมีวิธีการวิเคราะห์ซึ่งแบ่งได้เป็น 2 ประเภท คือ

1. การวิเคราะห์ตัวประกอบแบบ Component analysis นั้น จะไม่มีการจำแนกตัวประกอบร่วม (Common factors) จากตัวประกอบพิเศษ (Unique factors)

1.1 การวิเคราะห์ตัวประกอบหลัก (Principal component analysis)นี้ ถูกนำเข้ามาใช้ในการวิเคราะห์ตัวประกอบ โดย โฮลเทลลิ่ง (Hotelling, 1933)

2. การวิเคราะห์ตัวประกอบร่วม (Common factor analysis) นั้น จะถูกแบ่งตามวิธีการสะกัดตัวประกอบที่แตกต่างกันได้ 4 วิธี คือ

2.1 การวิเคราะห์แบบภาพพจน์ หรือการวิเคราะห์แบบเงา (Image analysis) ซึ่งพัฒนาโดย กัทแมน (Guttman, 1953)

2.2 การวิเคราะห์ตัวประกอบแบบคาโนนิคอล (Canonial factor analysis) ซึ่งพัฒนาโดย ราวและแฮร์ริส (Roa, 1955) และ (Harris, 1962)

2.3 การวิเคราะห์ตัวประกอบแบบแอลฟ่า (Alpha factor analysis) วิธีการนี้ถูกพัฒนาโดย ไกเซอร์และเคฟเฟรย์ (Kaiser and Caffrey, 1965)

2.4 วิธีการน้อยที่สุด (Minres method) นั้น จะถูกพัฒนาโดย ฮาร์แมนและ โจนส์ (Harman and Jones, 1966)

การวิเคราะห์ตัวประกอบแบบ Component analysis และการวิเคราะห์ตัวประกอบแบบ Common factor analysis จะมีความแตกต่างกันที่สำคัญ นั่นคือ จำนวนตัวประกอบและรูปแบบของการจำแนกความแปรปรวนของแต่ละ X_j (the j_{th} variable) โดยที่ในการวิเคราะห์ตัวประกอบแบบ Common factor analysis นั้น จะต้องมีข้อตกลงเบื้องต้นเกี่ยวกับค่าความร่วมกัน (Communality หรือ h_j^2) ซึ่งก็มีความหมายว่าตัวแปรนั้นมีค่าความแปรปรวนขนาดใดที่อธิบายได้จากจำนวนตัวประกอบหลัก ซึ่งในการวิเคราะห์นั้น รูปแบบการวิเคราะห์แบบ Common factor analysis จะต้องประมาณค่าความร่วมกันสำหรับแนวทะแยงในเมตริกซ์ที่จะวิเคราะห์กันก่อน อีกประการหนึ่งตัวประกอบในการวิเคราะห์ตัวประกอบแบบ

Common factor analyis นั้น จะสามารถอธิบายความสัมพันธ์ระหว่างตัวแปร โดยที่จำนวนตัวประกอบที่ได้จะมีจำนวนน้อยกว่าจำนวนตัวแปร ซึ่งจากที่กล่าวมานี้จะเห็นว่าความแปรปรวนทั้งหมดของตัวแปรแต่ละตัวนั้นจะถูกแบ่งออกเป็น 2 ส่วนย่อยๆ คือ ความแปรปรวนที่ร่วมกับตัวแปรอื่น ๆ (Common variance) และความแปรปรวนเฉพาะ (Unique variance) ในขณะที่การวิเคราะห์ตัวประกอบแบบ Component analysis นั้น จะมีจำนวนตัวประกอบเท่ากับจำนวนตัวแปร โดยที่จะสามารถอธิบายได้ว่าตัวประกอบนั้นสามารถอธิบายความแปรปรวนของ X_j ได้ทั้งหมด นั่นคือความแปรปรวนของแต่ละตัวแปรนั้นจะถูกแบ่งให้กับตัวประกอบทั้งหมดในชุดนั้น โดยที่จะเป็นการผสมกันระหว่างความแปรปรวนร่วมกันกับตัวแปรตัวอื่นและความแปรปรวนเฉพาะ

ในปัจจุบันนี้การวิเคราะห์ตัวประกอบแบบ Exploratary (Exploratory factor analysis) นี้ ยังคงเป็นวิธีการที่ได้รับความนิยมและใช้กันอยู่ในปัจจุบัน ประกอบกับที่คอมพิวเตอร์ได้เข้ามามีบทบาทในการวิเคราะห์ข้อมูล โดยที่ผู้วิจัยจะอาศัยโปรแกรมสำเร็จรูป (Computer - program package) มาช่วยในการวิเคราะห์ตัวประกอบได้สะดวกยิ่งขึ้น ยกตัวอย่างเช่น โปรแกรม SPSS SAS และ BMDP ซึ่งจะทำให้การวิเคราะห์ข้อมูลสะดวกและรวดเร็วยิ่งขึ้น และวิธีการวิเคราะห์ตัวประกอบที่ได้รับความนิยมมากที่สุดในอดีต และยังคงได้รับความนิยมกันอยู่ ในปัจจุบันนี้ก็คือการวิเคราะห์ตัวประกอบหลัก (Principal component analysis หรือ PC) นั้นเอง ซึ่งฮาร์แมน (Harman, 1967) ได้กล่าวไว้ว่าการวิเคราะห์ตัวประกอบหลักนี้ จะเป็นการสะกัดความแปรปรวนที่สูงที่สุดจากตัวแปรที่ถูกสังเกต (Observed Variables)โดยวิธีการนี้จะอาศัยหลักของความสัมพันธ์เชิงเส้นระหว่างตัวแปรที่ใช้เป็นข้อมูลทั้งหมด จะเป็นการอธิบายตัวประกอบโดย Linear combination ของตัวแปรดั้งเดิม (Original Variables) ที่สามารถอธิบายการผันแปรของข้อมูลได้มากที่สุดเท่านั้น แต่ไม่สามารถบอกได้ว่าตัวแปรตัวใดที่มีความสำคัญต่อตัวประกอบนั้นมากที่สุด และจำนวนตัวประกอบภายหลังการสะกัดนี้จะมีจำนวนเท่ากับจำนวนตัวแปรที่นำเข้า เช่นเดียวกันกับการวิเคราะห์ตัวประกอบแบบComponent analysis ซึ่งผู้วิจัยบางคนได้นำการวิเคราะห์ตัวประกอบหลักมาใช้ในงานวิจัย โดยมีความเข้าใจที่ว่าการวิเคราะห์ตัวประกอบแบบ Component analysis นี้ถือว่าเป็นการวิเคราะห์ตัวประกอบ (Factor analysis) ซึ่งเป็นความเข้าใจที่คลาดเคลื่อนกับความเป็นจริงที่ว่า การวิเคราะห์ตัวประกอบนั้นจะเป็นเทคนิคทางสถิติที่มีวัตถุประสงค์ในการลดปริมาณของตัวแปรให้มีจำนวนน้อยลง โดยที่จะสามารถค้นหาชุดของตัวแปรที่สามารถอธิบายได้ในรูปของตัวประกอบที่มีจำนวนน้อยกว่าจำนวนของตัวแปร พร้อมทั้งกำหนดว่าอะไรเป็นตัวประกอบที่เป็นตัวแทนของแต่ละคุณลักษณะได้ นอกจากนี้แล้วยังจะต้องมีทฤษฎีมารองรับอีก

ด้วย จากที่กล่าวมาข้างต้นนี้พอสรุปได้ว่า การวิเคราะห์ตัวประกอบแบบ Component analysis นี้ ไม่จัดว่าเป็นการวิเคราะห์ตัวประกอบ (Factor analysis) อย่างแน่นอน แต่สำหรับการวิเคราะห์ตัวประกอบที่ได้เข้ามามีบทบาทที่สำคัญ ในขณะนี้จะเป็นการวิเคราะห์ตัวประกอบแบบ Common factor analysis ซึ่งจะมีลักษณะเด่นหลายประการที่เป็นประโยชน์ต่องานวิจัยทางสังคมศาสตร์ ดังเช่นในด้านการวัดและการประเมินผลนั้นจะใช้การวิเคราะห์ตัวประกอบในการหาค่าความตรงของแบบสอบ และใช้ในการวิเคราะห์ข้อสอบโดยที่การวิเคราะห์นี้จะเป็นการวิเคราะห์ถึงความสัมพันธ์ระหว่างตัวแปรจำนวนมาก อีกทั้งจะมีการค้นหาชุดของตัวแปรที่สามารถอธิบายในรูปตัวประกอบได้ จะพิจารณาได้จากจำนวนตัวประกอบที่ถูกสะกัด และน้ำหนักตัวประกอบภายหลังจากการหมุนแกน โดยที่จำนวนของตัวประกอบที่ได้ภายหลังจากการสะกัดมีจำนวนตัวประกอบน้อยนั้น ย่อมแสดงให้เห็นว่าค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรเหล่านั้นว่ามีความสัมพันธ์กันสูง ทั้งนี้เนื่องจากว่าตัวแปรเหล่านั้นมีตัวประกอบร่วมกัน (Common factors) นอกจากนี้ยังจะสังเกตได้จากการจับกลุ่มของตัวแปรซึ่งเกิดขึ้นจากความสัมพันธ์ระหว่างกันและกันของตัวแปร ในทางกลับกันนั้นถ้าน้ำหนักตัวประกอบภายหลังการหมุนแล้ว ตัวแปรนั้นยังมีค่าน้ำหนักตัวประกอบมาก ย่อมจะแสดงว่าค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรและตัวประกอบร่วมที่ได้นั้น จะมีความสามารถในการอธิบายการจับกลุ่มของตัวประกอบที่ได้มากที่สุด นอกจากนี้ยังจะต้องพิจารณาถึงค่าความร่วมกันที่ได้จากการวิเคราะห์ตัวประกอบนี้เพื่อที่จะใช้ประโยชน์ในการแปลความหมายในการวิเคราะห์ตัวประกอบได้ชัดเจนยิ่งขึ้น

ลักษณะของข้อมูลที่จะนำมาใช้ในการวิเคราะห์ตัวประกอบนั้น จะขึ้นอยู่กับเป้าหมายและรูปแบบของการวิเคราะห์ตัวประกอบ โดยที่รูปแบบที่ใช้กันส่วนใหญ่คือการวิเคราะห์ตัวประกอบ ที่เรียกว่า ประเภท R (R - Type factor analysis) ในเรื่องระดับของตัวแปรที่จะนำมาใช้ในการวิเคราะห์ตัวประกอบนั้น จะต้องมีระดับการวัดแบบช่วง (Interval scale) หรืออาจจะเป็นตัวแปรทวิวิภาค (Dichotomous) ที่มีค่าเป็น 0 หรือ 1 เท่านั้น ในกรณีที่ผู้วิจัยมีจำนวนตัวแปร จำนวนตัวอย่าง หรือจำนวนประชากรมาก การใช้ข้อมูลดิบที่เป็นตัวแปรในการวิเคราะห์ตัวประกอบนั้น จะสิ้นเปลืองทั้งในด้านเวลา แรงงาน และ ค่าใช้จ่าย ดังนั้นผู้วิจัยควรหันมาใช้เมตริกซ์ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปร (Matrix of simple correlation coefficients) แทนข้อมูลในตัวแปร โดยค่าของการวิเคราะห์ตัวประกอบที่ได้จะสามารถคำนวณจากเมตริกซ์สหสัมพันธ์ ซึ่ง elements ในเมตริกซ์นี้ จะถูกคำนวณด้วยค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สัน โพรดักซ์ โมเมนต์ (Pearson's Product Moment Correlation Coefficient หรือ r_{xy}) โดยมีข้อตกลงเบื้องต้นเกี่ยวกับการใช้ค่าสัมประสิทธิ์นี้ในการคำนวณที่ว่า ข้อมูลในตัวแปรทั้งสอง

อย่างน้อยจะต้องจัดอยู่ในมาตราอันตรภาคชั้น(Interval scale) หรือมาตราอัตราส่วน (Ratio scale) โดยตัวแปรคู่นี้จะต้องเป็นตัวแปรที่มีค่าต่อเนื่อง (Continuous variables)

นอกจากนี้แล้วตัวแปรทั้งสองจะต้องมีความสัมพันธ์กันในเชิงเส้นตรงและมีการแจกแจงเป็นแบบปกติสองตัวแปร อีกทั้งจะต้องมีค่าความแปรปรวนที่เป็นเอกพันธ์ นั่นคือ การกระจายของตัวแปรตัวหนึ่งที่อยู่บนแกน X และตัวแปรอีกตัวหนึ่งอยู่บนแกน Y จะเป็นไปอย่างเดียวกัน โดยที่การแจกแจงของตัวแปรทั้งคู่นั้นจะต้องต่อเนื่องและสมมาตร แต่ไม่ถึงกับต้องเป็นการแจกแจงปกติ เช่นอาจจะมีฐานนิยม (Mode) เดียวกันในการแจกแจงก็ได้ ถ้าตัวแปรตัวหนึ่งเบ้ไปก็จะทำให้ค่าสัมประสิทธิ์สหสัมพันธ์นี้ลดลง (Wiersma, 1969) ซึ่งถ้าผู้วิจัยนั้นไม่สามารถทำตามข้อตกลงที่กำหนดไว้ข้างต้นนี้ ก็จำเป็นจะต้องใช้เทคนิค หรือ วิธีการอื่นๆในการคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์ โดยที่จะต้องคำนึงให้เหมาะสมกับระดับของข้อมูลที่มีอยู่ มิฉะนั้นจะทำให้ผลสรุปที่ได้นั้นไม่น่าเชื่อถือเท่าที่ควร ทั้งในด้านความแม่นยำและความตรงของการวิจัยนั้นด้วย

สำหรับกรณีของการวิเคราะห์ข้อสอบ ซึ่งมีการตรวจให้คะแนนสอบอยู่ในรูป 0, 1 โดยจะให้คะแนนเป็น 1 เมื่อตอบข้อสอบข้อนั้นได้ถูกต้อง และจะให้คะแนนเป็น 0 เมื่อตอบข้อสอบนั้นผิด โดยจะถือว่าคะแนนที่ได้จากการสอบนั้น เป็นตัวแปรที่ไม่ต่อเนื่องและได้จัดให้คะแนนดังกล่าวนี้อยู่ในมาตราการวัดแบบนามบัญญัติ (Nominal scale) ในการหาค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างข้อสอบของแบบสอบชุดนี้ ผู้วิเคราะห์อาจจะใช้ค่าสัมประสิทธิ์สหสัมพันธ์แบบฟี (Phi correlation coefficient) เป็นฐานสำหรับการวิเคราะห์ตัวประกอบ ทั้งนี้เนื่องจากว่าค่าสัมประสิทธิ์สหสัมพันธ์แบบฟีนั้น เป็นการวัดความสัมพันธ์ระหว่างตัวแปรทั้งสองตัวที่เป็นตัวแปรทวิวิภาคทั้งคู่ (Dichotomous) หรือตัวแปรตัวหนึ่งเป็นตัวแปรทวิวิภาค (Dichotomous) และอีกตัวหนึ่งนั้นอาจจะเป็นตัวแปรทวิวิภาคภายใต้การแจกแจงแบบปกติ (Artificial dichotomies with underlying normal distributions) เมเรนจ์ (Mehrens, 1967) ได้กล่าวว่าค่าสัมประสิทธิ์สหสัมพันธ์แบบฟีนั้น คือ อนุพันธ์ (Derivative) ของค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สัน โพรดักซ์ โมเมนต์ (Pearson's Product Moment Coefficient) ซึ่งสอดคล้องกับงานวิจัยของกิลฟอร์ดและเพอร์รี่ (Guilford and Perry, 1951) ที่ว่าค่าสัมประสิทธิ์สหสัมพันธ์แบบฟีนั้นจะเป็นการประมาณค่าของค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สัน โพรดักซ์ โมเมนต์ ส่วนกิลฟอร์ดและฟรัคเตอร์ (Guilford and Fruchter, 1978) ได้กล่าวไว้ว่า ค่าสัมประสิทธิ์สหสัมพันธ์แบบฟีนี้จะมีความสัมพันธ์ใกล้ชิดกับไคสแควร์ (Chi - square) ซึ่งสามารถที่จะนำมา

ประยุกต์ใช้ในกรณีต่างๆได้อย่างกว้างขวาง อีกทั้งค่าสัมประสิทธิ์สหสัมพันธ์นี้จะมีการแจกแจงแบบไม่ต่อเนื่อง (Discrete distribution) ถ้าข้อมูลที่นำเข้าไปในเมตริกซ์สหสัมพันธ์นั้นเป็นตัวแปรทวิวิภาคนั้น คือ จะมีค่าเป็น 0 หรือ 1 เท่านั้น ซึ่งถ้าดูโปรแกรมที่อยู่ในเครื่องคอมพิวเตอร์ส่วนใหญ่จะใช้ค่าสัมประสิทธิ์สหสัมพันธ์แบบฟีในการคำนวณเมตริกซ์สหสัมพันธ์ (SPSS Inc, 1988) ทั้งนี้เนื่องจากสูตรที่ใช้ในการคำนวณนั้นจะง่ายและสะดวกกว่าการใช้ค่าสัมประสิทธิ์สหสัมพันธ์แบบเตตระคลอริค จึงทำให้ค่าสัมประสิทธิ์สหสัมพันธ์แบบฟีได้รับใช้ในงานการวิเคราะห์ตัวประกอบจากความนิยมจากผู้วิจัยเป็นจำนวนมาก แต่ในปัจจุบันนี้ได้มีผู้ให้ความสนใจต่อทฤษฎีการตอบสนองรายข้อ (Item response theory) ทั้งนี้เนื่องมาจากทฤษฎีนี้เป็นการประเมินความสามารถที่แท้จริง (θ) ของผู้เข้าสอบนั้นไม่สามารถสังเกตได้โดยตรง แต่จะสามารถประมาณค่าความสามารถของผู้เข้าสอบที่ได้จากคะแนนที่สังเกตได้ (Observed score) จากการทำแบบทดสอบเช่นเดียวกันกับค่าความน่าจะเป็นของการตอบข้อสอบข้อที่ i ได้ถูกต้องของข้อสอบที่มีความสามารถระดับ θ $[P_i(\theta)]$ นั้น จะไม่สามารถสังเกตได้โดยตรงเช่นกันแต่สามารถจะประมาณค่าที่ได้จากสัดส่วนของผู้เข้าสอบที่ทำข้อสอบข้อที่ i ได้ถูกต้อง ซึ่งความสัมพันธ์ดังกล่าวสามารถแสดงในลักษณะโมเดลทางคณิตศาสตร์ที่กำหนดได้ด้วยฟังก์ชั่นทางคณิตศาสตร์ที่เรียกว่าโค้งคุณลักษณะของข้อสอบ (Item characteristic curve : ICC) โดยที่เป็นฟังก์ชั่นการถดถอยที่ไม่เป็นเชิงเส้นตรง (Non - linear regression function) ของคะแนนข้อสอบบนระดับความสามารถที่วัดโดยแบบสอบ จากที่กล่าวมาข้างต้นนี้จะเห็นได้ว่าความสามารถในการตอบข้อสอบโดยปกตินั้นจะมีความแตกต่างกันตั้งแต่ผู้เข้าสอบที่มีความสามารถต่ำที่สุด ไปยังผู้เข้าสอบที่มีความสามารถที่สูงที่สุดในประชากร โดยที่จะมีข้อตกลงเบื้องต้นว่า การแจกแจงของข้อสอบนั้นจะต้องมีความต่อเนื่อง และมีการแจกแจงแบบปกติ ซึ่งจะสอดคล้องกับ กิลฟอร์ดและฟรัคเตอร์ (Guilford and Fruchter, 1978) ที่ได้กล่าวไว้ว่า ค่าสัมประสิทธิ์สหสัมพันธ์ที่เหมาะสมกับตัวแปรทวิวิภาคภายใต้การแจกแจงแบบปกติ (Artificial dichotomous with underlying normal distribution) จะสามารถทำการประมาณค่าจากความสัมพันธ์เชิงเส้นตรง (Linear correlation) โดยที่ตัวแปรทั้งสองตัวนี้เป็นตัวแปรต่อเนื่อง (Continous variable) และมีการแจกแจงเป็นแบบปกติ (Normally distribution) และในงานวิจัยของกิลฟอร์ดและเพอร์รี่ (Guilford and Perry, 1951)นั้น สามารถที่จะคำนวณได้ทั้งค่าสัมประสิทธิ์สหสัมพันธ์แบบฟีหรือแบบเตตระคลอริค โดยที่ค่าสัมประสิทธิ์สหสัมพันธ์แบบฟีนั้น จะใช้ในการหาสหสัมพันธ์ระหว่างข้อสอบ คือ ต้องการรู้ว่าสามารถทำนายความสำเร็จบนข้อสอบนี้จากความสำเร็จบนข้อสอบอื่น ๆ ได้อย่างไร แต่ถ้าใช้สำหรับการวิเคราะห์ตัวประกอบที่มีพื้นฐานบนความสัมพันธ์ภายในของข้อสอบนั้น จะต้องใช้ค่าสัมประสิทธิ์สหสัมพันธ์แบบเตตระคลอริคในการคำนวณ

ในขณะนี้ได้มีข้อถกเถียงกันว่าสถิติพื้นฐานที่ใช้ในการคำนวณเมตริกซ์สหสัมพันธ์ที่ใช้กันอยู่ทุกวันนี้นั้น ควรจะใช้ค่าสัมประสิทธิ์ตัวใดในการคำนวณระหว่างค่าสัมประสิทธิ์สหสัมพันธ์แบบฟี หรือ ค่าสัมประสิทธิ์สหสัมพันธ์แบบเตตระคลอริค ซึ่งลอร์คและโนวิก (Lord and Novick, 1968) และกิลฟอร์ด (Guilford, 1941)นั้น ได้สนับสนุนให้ใช้สูตรการคำนวณค่าสัมประสิทธิ์สหสัมพันธ์แบบเตตระคลอริคมากกว่าที่จะให้ใช้สูตรการคำนวณค่าสัมประสิทธิ์สหสัมพันธ์แบบฟี แต่การใช้สูตรการคำนวณค่าสัมประสิทธิ์สหสัมพันธ์แบบเตตระคลอริคนั้นทำได้ยาก ทั้งนี้เนื่องจากจะต้องทำการคำนวณที่ซับซ้อนในโปรแกรม $SPSS_x$ แต่จะสามารถใช้โปรแกรม BMDP 4 F ได้ แต่ยังไม่มีการนำมาใช้อย่างแพร่หลาย อีกทั้งข้อมูลที่นำมาใช้นี้จะต้องเป็นไปตามข้อตกลงเบื้องต้นของการใช้ค่าสัมประสิทธิ์สหสัมพันธ์แบบเตตระคลอริค โดยที่แฮมบิงตัน และ สวามินาทาน (Hambleton and Swaminathan, 1985) ได้กล่าวว่าตัวแปรทั้งสองนั้นจะต้องมีการแจกแจงแบบปกติ แต่สำหรับในโปรแกรมคอมพิวเตอร์ส่วนใหญ่นั้น จะมีการใช้ค่าสัมประสิทธิ์สหสัมพันธ์แบบฟีในการคำนวณเมตริกซ์สหสัมพันธ์ นอกจากนี้แล้ว แฮมเบอตันและสวามินาทาน (Hambleton and Swaminathan, 1985) เชื่อว่า การสะกัดตัวประกอบที่คำนวณจากการใช้ค่าสัมประสิทธ์สหสัมพันธ์แบบฟีมาใช้แทนการใช้สูตรในการคำนวณค่าสัมประสิทธิ์สหสัมพันธ์แบบเตตระคลอริคนั้น ทำให้จำนวนตัวประกอบที่คำนวณจากค่าสัมประสิทธิ์สหสัมพันธ์ฟีมีจำนวนมากเกินไป แต่สำหรับคอมเรย์ (Comrey, 1973) นั้นสนับสนุนให้ใช้ค่าสัมประสิทธิ์สหสัมพันธ์แบบฟีกับตัวแปรไดโคโตมัส (Dichotomous variables) อีกทั้งยังกล่าวว่าการใช้ค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สันนั้นเหมาะสมในกรณีทั่วไปมากกว่าถึงแม้ว่าในปัจจุบันนี้จะมีความคลาดเคลื่อนมากกว่าความเป็นจริง

จากผลการศึกษาและข้อเสนอแนะต่างๆจะพบว่า ยังไม่สามารถหาข้อสรุปที่มีหลักฐานที่ชัดเจนเพียงพอเกี่ยวกับการเลือกใช้ค่าสัมประสิทธิ์สหสัมพันธ์ในการคำนวณเมตริกซ์สหสัมพันธ์ที่เหมาะสม สำหรับการวิเคราะห์ตัวประกอบที่ใช้ค่าสัมประสิทธิ์สหสัมพันธ์แบบฟี กับแบบเตตระคลอริค เมื่อตัวแปรนั้นจะมีลักษณะเป็นไดโคโตมัส ปัญหาเบื้องต้นที่ควรได้รับการพิจารณาศึกษาเปรียบเทียบก็คือ เมื่อผู้วิเคราะห์ใช้เมตริกซ์สหสัมพันธ์ที่คำนวณจากค่าสัมประสิทธิ์สหสัมพันธ์แบบเตตระคลอริคในการวิเคราะห์ตัวประกอบนั้น ผลที่ได้จะแตกต่างจากการวิเคราะห์ตัวประกอบที่ใช้เมตริกซ์สหสัมพันธ์ที่คำนวณจากค่าสัมประสิทธิ์สหสัมพันธ์แบบฟีหรือไม่ หากผลการวิเคราะห์ไม่มีความแตกต่างกัน ผู้วิเคราะห์จะเลือกใช้วิธีใดก็น่าจะใช้ได้ แต่ถ้ามีความแตกต่างกันในผลการวิเคราะห์จึงควรจะได้พิจารณาให้ลึกซึ้งต่อไป

วัตถุประสงค์ของการวิจัย

การวิจัยครั้งนี้มุ่งเปรียบเทียบผลการวิเคราะห์ตัวประกอบระหว่างการใช้ค่าสัมประสิทธิ์สหสัมพันธ์แบบฟี กับการใช้ค่าสัมประสิทธิ์สหสัมพันธ์แบบเตตระคลอริคในการคำนวณค่าในเมตริกซ์สหสัมพันธ์ (Matrix of Correlation) โดยจะพิจารณาเปรียบเทียบผลการวิเคราะห์ เมื่อใช้ความยาวของแบบสอบ และขนาดของกลุ่มตัวอย่างที่แตกต่างกัน ตามหัวข้อต่อไปนี้คือ

1. ค่าความร่วมกัน (Communality)
2. น้ำหนักตัวประกอบหลังจากการหมุนแกนแล้ว (Factor loading)
3. จำนวนตัวประกอบ

สมมติฐานการวิจัย

จากการศึกษาของแฮมเบอตันและสวามินาทาน (Hambleton and Swaminathan, 1985) มีความเชื่อที่ว่าการใช้ค่าสัมประสิทธิ์สหสัมพันธ์แบบฟีนั้น จะทำให้มีจำนวนตัวประกอบที่มากกว่าการใช้ค่าสัมประสิทธิ์สหสัมพันธ์แบบเตตระคลอริค และ จากการศึกษาของโคเฮน(Cohen, 1983) ได้กล่าวว่า ค่าสัมประสิทธิ์สหสัมพันธ์เตตระคลอริคนั้นจะมีค่าใหญ่กว่าค่าสัมประสิทธิ์สหสัมพันธ์แบบฟี ดังนั้นผู้วิจัยจึงตั้งสมมติฐานการวิจัยดังนี้ คือ

1. ค่าความร่วมกันที่ได้จากการวิเคราะห์ตัวประกอบโดยใช้เมตริกซ์สหสัมพันธ์ที่คำนวณได้จากค่าสัมประสิทธิ์สหสัมพันธ์แบบเตตระคลอริคนั้น น่าจะแตกต่างจากค่าความร่วมกันที่ได้จากการวิเคราะห์ตัวประกอบโดยใช้เมตริกซ์สหสัมพันธ์ที่คำนวณได้จากค่าสัมประสิทธิ์สหสัมพันธ์แบบฟี

2. การวิเคราะห์ตัวประกอบโดยใช้เมตริกซ์สหสัมพันธ์ที่คำนวณจากค่าสัมประสิทธิ์สหสัมพันธ์แบบเตตระคลอริคนั้น จะทำให้น้ำหนักตัวประกอบภายหลังการหมุนแกนแล้วมีน้ำหนักแตกต่างจากการวิเคราะห์ตัวประกอบโดยใช้เมตริกซ์สหสัมพันธ์ที่คำนวณจากค่าสัมประสิทธิ์สหสัมพันธ์แบบฟี

ขอบเขตของการวิจัย

1. ในการศึกษาเปรียบเทียบนี้ จะมุ่งเปรียบเทียบเฉพาะผลการวิเคราะห์ตัวประกอบระหว่างการใช้ค่าสัมประสิทธิ์สหสัมพันธ์แบบฟี กับการใช้ค่าสัมประสิทธิ์สหสัมพันธ์แบบเตตระคลอริค ในการคำนวณค่าในเมตริกซ์สหสัมพันธ์ (Matrix of correlation) โดยศึกษาตัวแปรต่อไปนี้

1.1 ตัวแปรตาม (Dependent variables) ได้แก่

ก. ค่าความร่วมกัน (Communality)

ข. น้ำหนักตัวประกอบหลังจากการหมุนแกนแล้ว (Factor loading)

ค. จำนวนตัวประกอบ

1.2 ตัวแปรอิสระ (Independent variables) ได้แก่

ก. เมตริกซ์สหสัมพันธ์ที่คำนวณได้จากค่าสัมประสิทธิ์สหสัมพันธ์แบบฟี และเมตริกซ์สหสัมพันธ์ที่คำนวณได้จากค่าสัมประสิทธิ์สหสัมพันธ์แบบเตตระคลอริค

ข. ความยาวของแบบสอบ

ค. ขนาดของกลุ่มตัวอย่าง

2. ศึกษาเฉพาะกรณีที่มีจำนวนของกลุ่มตัวอย่างเพียงกลุ่มเดียวเท่านั้น โดยที่การสุ่มข้อสอบจำนวน 20 ข้อจากจำนวนข้อสอบทั้งหมด 94 ข้อ โดยจะทำการสุ่มจำนวนของผู้เข้าสอบให้มีจำนวนดังนี้คือ 400 คน, 600 คน, 800 คน, 1,000 คน, 1,200 คน, 1,400 คน และ 1,600 คน ตามลำดับ จากนั้นสุ่มข้อสอบให้มีจำนวน 30 ข้อ, 40 ข้อ, 50 ข้อ, 60 ข้อ, 70 ข้อ และ 80 ข้อ โดยผู้วิจัยจะให้จำนวนข้อสอบชุดเดียวกันในทุกขนาดของกลุ่มตัวอย่าง แต่จะทำการสุ่มจำนวนผู้เข้าสอบใหม่ทุกครั้ง

3. ในแต่ละกรณีของการทดลองครั้งนี้ จะทำการทดลองซ้ำ 100 ครั้ง

ข้อจำกัดของการวิจัย

1. เนื่องจากการวิจัยครั้งนี้ได้กำหนดแผนการเปรียบเทียบผลการวิเคราะห์ไว้ถึง 49 กรณี ดังนั้นในแต่ละละกรณีจึงใช้วิธีการทดลองซ้ำเพียง 100 ครั้ง

2. สำหรับการเปรียบเทียบค่าเฉลี่ยของน้ำหนักตัวประกอบภายหลังจากการหมุนแกนที่วิเคราะห์จากเมตริกซ์สหสัมพันธ์แบบฟีกับแบบเตตระคลอริคนั้น จะเลือกเฉพาะตัวประกอบแรก (First factor) เท่านั้น

3. ในการวิเคราะห์ตัวประกอบที่ใช้ในงานวิจัยครั้งนี้ จะใช้วิธีการสะกัดตัวประกอบแบบภาพพจน์ (Image analysis) และหมุนแกนตัวประกอบด้วยวิธีวาริแมกซ์ (Varimax rotation) เท่านั้น

นิยามศัพท์ที่ใช้ในการวิจัย

การวิเคราะห์ตัวประกอบ (Factor Analysis) หมายถึง เทคนิคทางสถิติสำหรับสะกัดตัวประกอบพื้นฐานภายใต้จำนวนตัวแปรที่มากกว่า โดยจะเริ่มต้นวิเคราะห์จากสหสัมพันธ์ระหว่างตัวแปรเหล่านั้น ซึ่งมีวัตถุประสงค์ในการลดปริมาณของตัวแปรให้มีจำนวนน้อยลง โดยอาศัยโครงสร้างและแบบแผนของความสัมพันธ์ที่มีระหว่างตัวแปร ทั้งนี้เพื่อให้ง่ายต่อความเข้าใจ และทำให้ทราบถึงโครงสร้างและแบบแผนของข้อมูลนั้น

ผลการวิเคราะห์ตัวประกอบ หมายถึง การพิจารณาค่าที่ได้จากการวิเคราะห์ตัวประกอบดังนี้คือ ค่าความร่วมกัน(Communality) น้ำหนักตัวประกอบภายหลังจากการหมุนตัวประกอบแล้ว (Factor loading) และจำนวนตัวประกอบ

ค่าความร่วมกันของตัวแปร (Communality) หมายถึง สำหรับตัวแปรในการวิเคราะห์ตัวประกอบ ค่าความร่วมกันจะเป็นผลบวกกำลังสอง (Sum of Square) ของน้ำหนักตัวประกอบ (Factor loading) ทั้งหมดของตัวประกอบ ค่าที่ให้นี้จะเป็นดัชนีบ่งชี้ว่าตัวแปรนั้นซ้อนกัน (Overlap)กับตัวประกอบหรือเท่าเทียมกัน ซึ่งสามารถถูกนับ (Accounted by)คะแนนในตัวประกอบ ในหลาย ๆ วิธีของการสะกัดตัวประกอบนั้น ในการประมาณค่าของค่าความร่วมกันแทนที่เซลไดออกอนอล (Diagonal cells) ของเมตริกซ์สหสัมพันธ์ดั้งเดิม (Original correlation matrix) ก่อนการวิเคราะห์ตัวประกอบจะสำเร็จ (Carried out) หรืออาจจะกล่าวได้ว่า

ตัวแปรตัวนั้นมีค่าความแปรปรวนขนาดใดที่อธิบายได้จากจำนวนตัวประกอบหลัก ซึ่งถ้าเป็นการวิเคราะห์แบบ Component analysis จะแทนค่าในแนวทะแยงของเมตริกซ์เท่ากับ 1 ในขณะที่การวิเคราะห์แบบ Common factor analysis จะต้องประมาณค่าความร่วมกันสำหรับแนวทะแยงในเมตริกซ์ก่อน ซึ่งสามารถประมาณค่าความร่วมกันของตัวแปร โดยใช้ Square Multiple Correlation (SMC_j) นั่นคือ ค่ากำลังสองของสหสัมพันธ์พหุคูณของตัวแปรตัวที่ j กับตัวแปรที่เหลือ

สหสัมพันธ์ (Correlation) หมายถึง ค่าวัดที่ใช้วัดความมากน้อยในความสัมพันธ์เชิงเส้นตรงระหว่างตัวแปรสุ่มทั้งสอง

น้ำหนักขององค์ประกอบ (Factor Loading : a_{ij}) หมายถึง ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวประกอบร่วมกับตัวแปรหรืออาจจะกล่าวได้ว่าเป็นความสัมพันธ์ของตัวแปรตัวนั้น กับตัวประกอบตัวนั้น

สัมประสิทธิ์สหสัมพันธ์แบบฟี (Phi correlation coefficient : r_ϕ) หมายถึง สัมประสิทธิ์สหสัมพันธ์ที่ถูกใช้เป็นดัชนีบ่งบอกความสัมพันธ์ระหว่างตัวแปรทวิวิภาค (Dichotomous)สองตัว โดยที่ตัวแปรนั้นจะมีค่าเป็น 0 หรือ 1 ค่าสัมประสิทธิ์สหสัมพันธ์นั้นมีค่าอยู่ระหว่าง -1 ถึง 1 โดยที่ค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนบนข้อสอบทั้งสอง จะเป็นประโยชน์ในโครงสร้างของแบบทดสอบ โดยที่จะมีสูตรที่นิยมใช้ในปัจจุบันนี้จะเป็นการคำนวณจากความถี่ (Freqency) ดังนี้คือ

$$r_\phi = \frac{bc - ad}{\sqrt{(a+b)(a+c)(b+d)(c+d)}}$$

(Lindeman and et al., 1980)

โดยที่ a, b, c และ d เป็นความถี่ในแต่ละเซลที่จะต้องมีความถี่ตรงกับที่แสดงในตารางดังนี้คือ

		ตัวแปรที่ 1	
		0	1
ตัวแปรที่ 2	1	a	b
	0	c	d

สัมประสิทธิ์สหสัมพันธ์เตตระคลอริค (Tetrachoric correlation : r_{tet}) หมายถึง สหสัมพันธ์ของตัวแปรที่เป็นตัวแปรทวิวิภาคภายใต้การแจกแจงปกติทั้งคู่ (Artificial dichotomies with underlying normal distribution) โดยมีข้อตกลงเบื้องต้นเกี่ยวกับตัวแปรทั้งคู่นั้น จะต้องต่อเนื่อง (Continuous) มีการแจกแจงเป็นปกติ (Bivariate normal distribution) และ ตัวแปรเหล่านี้จะมีความสัมพันธ์เชิงเส้นตรง (Linearly related variable) ค่าสัมประสิทธิ์สหสัมพันธ์แบบเตตระคลอริคนี้ จะใช้วัดความสัมพันธ์ระหว่างคะแนนบนแบบสอบรายข้อ โดยทั่วไปนี้จะมีสูตรที่คำนวณจากความถี่ที่ได้รับจากตารางการณ์จร ซึ่งนับว่าเป็นสูตรที่อยู่ในรูปแบบที่ง่ายดังสูตรที่ปรากฏ ดังนี้คือ

$$r_{tet} = \cos\left(\frac{180^{0}\ \sqrt{bc}}{\sqrt{ad}\ +\ \sqrt{bc}}\right)$$

(Guilford and Fruchter, 1978)

โดยที่ a, b, c และ d เป็นความถี่ในแต่ละเซลที่จะต้องมีความถี่ตรงกับที่แสดงในตารางดังนี้คือ

		ตัวแปรที่ 1	
		0	1
ตัวแปรที่ 2	1	b	a
	0	d	c

ซึ่งสูตรที่แสดงข้างบนนี้ จะเป็นเพียงค่าประมาณของค่าสัมประสิทธิ์สหสัมพันธ์แบบเตตระคลอริคเท่านั้น

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ประโยชน์ที่คาดว่าจะได้รับ

ผลการวิจัยนี้จะช่วยให้ผู้ใช้สถิติ มีข้อสรุปและหลักฐานในการเลือกใช้วิธีการหาค่าสัมประสิทธิ์สหสัมพันธ์ เมื่อข้อมูลมีลักษณะเป็นไดโคโตมัสได้อย่างเหมาะสมกับสภาพการณ์ปัจจุบันของข้อมูล อันจะทำให้ผลสรุปหรือข้อความรู้ที่ได้จากงานวิจัยนั้น ๆ มีความเชื่อถือได้

1. ได้ทราบว่าวิธีที่เหมาะสมที่ใช้ในการวิเคราะห์ตัวประกอบจากเมตริกซ์สหสัมพันธ์ที่คำนวณจากค่าสัมประสิทธิ์สหสัมพันธ์แบบฟี กับแบบเตตระคลอริค เมื่อข้อมูลนั้นมีลักษณะเป็นตัวแปรทวิวิภาค
2. เพื่อให้นักวิจัยมีผลสรุป และ หลักฐานในการเลือกใช้วิธีการทางสถิติในการหาค่าสัมประสิทธิ์สหสัมพันธ์เกี่ยวกับข้อมูลที่มีลักษณะไดโคโตมัสได้อย่างเหมาะสม
3. เป็นหลักฐานในการวิจัยด้านนี้ต่อไป

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย