บทที่ 2

วรรณคดีที่เกี่ยวข้อง

ในส่วนของวรรณคดีที่เกี่ยวข้องนี้ ผู้วิจัยแบ่งออกเป็น 4 ตอน ดังนี้คือ

ตอนที่ 1 มาตราการวัด
ตอนที่ 2 การวิเคราะห์ตัวประกอบ
ตอนที่ 3 สัมประสิทธิ์สหสัมพันธ์ที่ศึกษาในงานวิจัยนี้
ตอนที่ 4 งานวิจัยที่เกี่ยวข้อง

ตอนที่ 1 มาตราการวัด (Measurement Scales)

ในการใช้สถิติเพื่อวิเคราะห์ข้อมูลในการวิจัยนั้น ผู้วิจัยจำเป็นต้องทราบลักษณะของข้อมูลที่ใช้ในการวิจัยเสียก่อน เพื่อเป็นเครื่องวินิจฉัยว่าจะใช้วิธีการทางสถิติแบบใดในการวิเคราะห์จึงจะถูกต้องเหมาะสม นอกจากนี้แล้วผู้วิจัยยังจะต้องทำความเข้าใจเกี่ยวกับตัวแปรโดยที่มาตราต่างๆ ที่ใช้วัดค่าของแต่ละตัวแปรนั้น นับว่าเป็นสิ่งสำคัญที่ควรได้ทำความเข้าใจให้ชัดเจนในขั้นต้น เพราะการแปรค่าของตัวแปรไม่ว่าจะเป็นตัวแปรใดก็ตาม จะแปรค่าตามมาตราการวัด ดังนั้นผู้วิจัยจึงควรทราบถึงมาตราการวัดของข้อมูลก่อน โดย สตีเว่นส์ (Stevens, 1946) ได้จำแนกมาตราการวัดออกเป็น 4 ระดับ ดังนี้

1. มาตรานามบัญญัติ (Categorical หรือ Nominal scale) เป็นมาตราการวัดในระดับที่ต่ำสุด โดยที่ตัวเลขที่กำหนดขึ้นเป็นเพียงการจำแนกหรือแยกประเภท (Classification) ลักษณะต่างๆภายในตัวแปรที่แปรค่าตามมาตรานี้ ซึ่งอาจจะเป็นเพียงการเรียกชื่อเท่านั้น ดังนั้นการวัดในระดับนี้บางทีจึงไม่เป็นที่ยอมรับว่าเป็น " การวัด " เพราะไม่สามารถบอกปริมาณมากน้อยได้ เป็นแต่เพียงแสดงให้เห็นความแตกต่างของสิ่งต่าง ๆ เท่านั้น เช่น ศาสนาที่นับถือ อาจจะแปรค่าเป็นคนที่นับถือศาสนาพุทธ คริสต์ อิสลาม ถ้ากำหนดให้ศาสนาพุทธ = 1, ศาสนาคริสต์ =

2, ศาสนาอิสลาม = 3 ดังนั้นตัวเลข 1, 2 และ 3 จะบอกถึงศาสนาที่นับถือตามที่กำหนดไว้เท่านั้น แต่มิได้สื่อความหมายทางคณิตศาสตร์แต่ประการใด หรือ ตัวแปรเพศ ถ้ากำหนดให้ เพศชาย = 1 และ เพศหญิง = 0 แบบนี้ก็มิได้หมายความว่า 1 จะมากกว่า 0 ตัวอย่างเพิ่มเติมของตัวแปรประเภทนี้ เช่น ประเภทของที่อยู่อาศัย ที่ตั้งของโรงเรียน เป็นต้น ถ้าตัวแปรที่แปรตามมาตรานี้สามารถจำแนกลักษณะได้เพียง 2 ลักษณะ เรียกว่า Dichotomous variable เช่น เพศ (ชาย , หญิง) ผลการสอบ (ผ่าน , ไม่ผ่าน)เป็นต้น แต่ถ้าตัวแปรที่แปรตามมาตรานี้ สามารถจำแนกลักษณะได้มากกว่า 2 ลักษณะ เรียกว่า Polytomous variable เช่น การนับถือศาสนา, สถานภาพในการสมรส, ภูมิลำเนา, อาชีพ เป็นต้น

การกำหนดตัวเลขให้กับสิ่งต่างๆในมาตรานามบัญญัติ จะเป็นเพียงตัวแทนประเภทของสิ่งที่ถูกวัดหรือเพื่อใช้ในการสื่อความหมายให้จำได้ง่าย หรือสะดวกในการนับ โดยที่ตัวเลขดังกล่าวไม่มีความหมายในเชิงปริมาณแต่ประการใด ดังนั้นจึงไม่สามารถนำตัวเลขเหล่านั้นมาบวก ลบ คูณ หรือ หารกันได้ โดยเฉพาะทางด้านคุณภาพที่ไม่สามารถจำแนกถึงขนาดของความแตกต่างแน่นอนได้ว่ามีลักษณะ "มากกว่า" หรือ "น้อยกว่า" กัน คงชี้ได้แต่เพียงว่าตัวแปรหรือข้อมูลนั้นมีความแตกต่างกันเท่านั้น เพราะฉะนั้นเทคนิคและสูตรทางด้านสถิติที่จะนำมาใช้กับจำนวนที่วัดด้วยมาตรานี้ จะเป็นสูตรที่ใช้เพื่อช่วยในการการนับจำนวน หรือ หาความถี่ของลักษณะที่เหมือนกันเท่านั้น

คุณสมบัติที่สำคัญของมาตรานามบัญญัติคือ ความเท่าเทียมกัน (Equivalence) กล่าวคือ สมาชิกที่อยู่ในกลุ่มเดียวกันจะมีคุณค่าเหมือนกัน ตัวเลขหรือสัญลักษณ์ที่กำหนดให้เป็นเพียงชื่อ ไม่อาจนำมาใช้ในการคำนวณทางคณิตศาสตร์ ตัวเลขหรือสัญลักษณ์เหล่านี้ถ้าได้กำหนดให้หมายถึงอะไรแล้ว จะต้องไม่ใช้ตัวเลขหรือสัญลักษณ์นั้นซ้ำกับสิ่งอื่นๆ อีกเป็นอันขาด และในทำนองเดียวกัน ผู้วิจัยไม่ควรใช้ตัวเลข 2 ตัวซ้ำกันไปใช้แทนสิ่งของชิ้นเดียวกันด้วย เพราะจะทำให้งง ไม่ทราบว่า ตัวเลขตัวใดที่แทนสิ่งของนั้น โดยสรุปแล้ว มาตรานามบัญญัตินี้จะมีคุณลักษณะดังต่อไปนี้

1.1 อยู่อย่างไม่เป็นระเบียบ ต่างคนต่างอยู่ หรือแยกกันอยู่

1.2 ข้อมูลที่วัดออกมาเป็นกลุ่มหรือประเภทต่าง ๆ ไม่จำเป็นต้องมีความสัมพันธ์กี่ยวข้องกัน

1.3 ในมาตรานามบัญญัติจะใช้ตัวเลข ตัวอักษร หรือ สัญลักษณ์เครื่องหมายอย่างใดก็ได้ ตัวเลขหรือสัญลักษณ์เหล่านี้ไม่มีความหมายในแง่ปริมาณ หรือคุณค่าที่จะนำไปเปรียบเทียบกันได้

1.4 สมาชิกที่อยู่ในกลุ่มเดียวกันจะมีคุณค่า หรือคุณสมบัติเหมือนกัน อาจจะสลับ

สับเปลี่ยนกัน หรือทดแทนกันได้

2. มาตราอันดับ (Ordinal scale) การวัดด้วยมาตราอันดับนี้ จะเป็นการวัดที่มีความละเอียด (Refinement) ของการวัดมากขึ้น เมื่อเทียบกับมาตรานามบัญญัติ ลักษณะการวัดเป็นการจัดอันดับ (Rank order) ให้กับคุณสมบัติบางอย่างว่ามากกว่าหรือน้อยกว่าข้อมูลอื่นๆ ซึ่งสามารถจัดอันดับข้อมูลตามตำแหน่งได้จากน้อยที่สุดไปหามากที่สุด หรือน้อยที่สุดไปหามากที่สุดได้ ตัวอย่างของตัวแปรที่แปรตามมาตรานี้ คือ ผลการแข่งขันกีฬามหาวิทยาลัย ระดับความคิดเห็น ตำแหน่งเปอร์เซนต์ไทล์ ตำแหน่งทางวิชาการ เช่น ศาสตราจารย์ รองศาสตราจารย์ ผู้ช่วยศาสตราจารย์ และ อาจารย์ ลำดับที่ผลการเรียนในชั้น ลำดับที่ของความแข็งของวัตถุเมื่อเทียบกับค่าอื่น ๆ ในกลุ่มเดียวกัน เช่น นักเรียนที่สอบได้ลำดับที่ 2 นั้น จะมีคะแนนอยู่ในตำแหน่งที่สูงกว่านักเรียนที่สอบได้ลำดับที่ 3 เพียงแต่ว่าไม่ได้วัดละเอียดลงไปถึงตัวค่าจริงของสิ่งที่จะวัดนั้น ถึงแม้ว่าค่าของสิ่งที่วัดจะเปลี่ยนไปแต่ถ้าตำแหน่งยังเหมือนเดิม ตัวเลขที่วัดได้ตามมาตรานี้ก็จะยังคงเหมือนเดิม นอกจากนี้แล้วผลการวัดในมาตรานี้ไม่อาจเปรียบเทียบปริมาณของความแตกต่างของแต่ละอันดับเหล่านั้นได้ว่าแตกต่างกันมากน้อยเพียงใด เช่น จะบอกว่านักเรียนที่สอบได้ที่ 1 เก่งกว่านักเรียนที่สอบได้ที่ 2 เท่ากับ นักเรียนที่สอบได้ที่ 3 เก่งกว่านักเรียนที่สอบได้ที่ 4 ไม่ได้ แต่จะสามารถบอกได้แต่เพียงว่าต่างกันไปในทางไหนหรือทำให้ทราบทิศทางเท่านั้น ดังนั้นการวัดในมาตรานี้จึงยังนับว่ายังวัดได้ไม่ละเอียดเท่าที่ควร แต่ตัวเลขที่วัดได้นั้นจะละเอียด และสื่อความหมายมากขึ้นเมื่อเทียบกับมาตรานามบัญญัติ

3. มาตราช่วง หรือ อันตรภาค (Interval scale) มาตรานี้จะมีการวัดค่าที่ละเอียดไปถึงค่าจริงของสิ่งที่วัด ตัวอย่างของตัวแปรที่แปรค่าตามมาตรานี้ เช่น คะแนนผลการเรียนในวิชาต่างๆ ปีปฏิทิน อุณหภูมิ เป็นต้น ตัวเลขที่วัดได้จากมาตรานี้ จะมีช่วงห่างที่เท่ากัน จึงสามารถเปรียบเทียบกันได้ว่ามากน้อยกว่ากันเท่าใด แต่ไม่สามารถบอกได้ว่าเป็นกี่เท่าของกันและกัน เพราะมาตรานี้ไม่มีศูนย์แท้หรือศูนย์สมบูรณ์ (True zero or Absolute zero) คือค่าที่เป็นศูนย์ในมาตรานี้เป็นเพียงค่าที่กำหนดขึ้นเท่านั้น เช่น นิสิตที่สอบได้คะแนน 0 ไม่ได้หมายความว่า เขาไม่มีความรู้เลยในวิชานั้น ดังนั้นมาตรานี้จึงสามารถเปรียบเทียบความแตกต่างของสิ่งที่วัดได้ แต่จะบอกแตกต่างเป็นจำนวนเท่าไม่ได้เช่นจะบอกได้ว่าความแตกต่างระหว่างอุณหภูมิ 50 °F กับ 60 °F จะเท่ากับความแตกต่างระหว่างอุณหภูมิ 90 °F กับ 100 °F แต่ไม่สามารถกล่าวได้ว่าอุณหภูมิ 100 °F จะมีความร้อนเป็น 2 เท่าของ 50 °F ทั้ง ๆ ที่ในทางคณิตศาสตร์นั้น 100 มีค่าเป็นสองเท่าของ 50 และที่กล่าวว่าอุณหภูมิ 0 °F หรือ 0 °C ไม่ได้

หมายความว่าอุณหภูมิ ณ จุดนี้ไม่มีความร้อนอยู่ มาตราอันตรภาคนี้สามารถเปลี่ยน (Convert) เป็นมาตราอันดับได้ง่าย แต่มาตราอันดับไม่สามารถแปลง (Transform) เป็นมาตราอันตรภาค ยกตัวอย่างเช่น จำนวนวันในเดือนกรกฎาคมนั้น สามารถจัดอันดับจากวันที่มีความร้อนมากที่สุดไปหาวันที่มีความร้อนน้อยที่สุด แต่จะไม่สามารถที่จะแปลงอันดับนี้เข้าไปในหน่วยที่ใช้การวัดอุณหภูมิที่เป็นฟาเรนไฮต์ (Degree fahrenheit) กลาสและฮอบกิ้นส์ (Glass and Hopkins, 1984) ได้กล่าวว่า มาตราอันดับ (Ordinal measurement) จะใช้ประโยชน์ได้มากกว่ามาตราช่วงหรือมาตราอัตราส่วน (Interval or Ratio measurement) เช่น มีเด็กผู้ชายอายุ 10 ปี มีความสูง 50 นิ้ว และมีน้ำหนัก 80 ปอนด์นี้ จะให้ข่าวสารที่น้อยกว่าที่จะบอกว่า เขามีความสูงและน้ำหนักอยู่ในเปอร์เซนต์ไทล์ที่ 5 และ เปอร์เซนต์ไทล์ที่ 95 ตามลำดับ โดยสรุปมาตราอันตรภาคนี้จะมีลักษณะดังต่อไปนี้คือ

3.1 สิ่งที่แบ่งหรือจัดประเภทในกลุ่มเดียวกัน จะห่างกันเป็นช่วงแต่ละช่วงที่มีค่าความห่างที่เท่ากัน

3.2 จุดเริ่มต้นของแต่ละประเภทนั้น ไม่มีจุดเริ่มต้นที่เป็นไปตามธรรมชาติ จุดเริ่มต้นมีฐานะเป็นเพียงศูนย์สัมพัทธ์ หรือศูนย์เทียม (Relative or Arbitary zero) เท่านั้น เช่นน้ำ 0 ˚C มิใช่ว่าน้ำนั้นไม่มีความร้อนเลย

3.3 อาจจะกำหนดตัวเลขแทนวัตถุ สิ่งของแต่ละอย่างนั้นได้ และความแตกต่างระหว่างตัวเลข จะแทนค่าความแตกต่างของวัตถุสิ่งของนั้นด้วย

4. มาตราอัตราส่วน (Ratio scale) เป็นมาตราการวัดในระดับที่สูงที่สุดที่จะใช้จัดข้อมูล และสามารถแสดงความสัมพันธ์ทางด้านตัวเลขได้ นอกจากนี้มาตรานี้มีลักษณะเพิ่มเติมจากมาตราอันตรภาค คือเป็นมาตราที่เริ่มวัดจากค่าศูนย์แท้หรือ ศูนย์สมบูรณ์ (Absolute zero) คือ ตัวเลขศูนย์บนมาตรานี้ไม่ว่าระบบใดที่ใช้วัดค่าข้อมูลชนิดเดียวกัน จะวัดมูลค่าของตัวแปรได้เท่ากัน ข้อมูลในมาตรานี้สามารถนำมาบวก ลบ คูณ หารกันได้ ตัวอย่างที่จัดอยู่ในมาตรานี้มักจะเป็นการวัดทางฟิสิกส์ เช่น ความสูง ความยาว น้ำหนัก เวลา พื้นที่ มุม เป็นต้น สำหรับในทางสังคมศาสตร์ส่วนใหญ่นั้น ยังไม่สามารถวัดได้ถึงมาตราอัตราส่วนนี้ เพราะไม่ทราบจุดเริ่มต้นที่แท้จริงของสิ่งที่ต้องการจะวัด โดยเฉพาะการวัดที่เกี่ยวกับพฤติกรรมของบุคคลในด้านต่าง ๆ

ตอนที่ 2 แนวคิดและทฤษฎีของการวิเคราะห์ตัวประกอบ

ประวัติการวิเคราะห์ตัวประกอบ

ในระยะเวลาประมาณ ค.ศ. 1900 หลังจากนักสถิติชาวอังกฤษ ชื่อ คาร์ล เพียร์สัน(Karl Pearson) ได้พัฒนาวิธีหาความสัมพันธ์ระหว่างตัวแปร โดยใช้ค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สัน โพรดักซ์ โมเมนต์ (Pearson Product Moment Correlation Coefficient : r_{xy}) แล้ว ปรากฏว่า มีผู้ค้นพบความสัมพันธ์ระหว่างตัวแปรต่างๆมากมาย จึงเกิดมีความต้องการที่จะอธิบายถึงโครงสร้างของความสัมพันธ์เหล่านั้น ได้มีนักจิตวิทยาชาวอังกฤษชื่อ ชาร์ล สเปียร์แมน (Charles Spearman) ซึ่งได้ชื่อว่าเป็นผู้ก่อตั้งทฤษฎีเกี่ยวกับเชาวน์ปัญญา (Intelligent) และยังได้ชื่อว่าเป็น "บิดาของการวิเคราะห์ตัวประกอบ" ได้เสนอโมเดลของการวิเคราะห์ตัวประกอบแบบแรก ๆ ไว้ในปี ค.ศ. 1904 จากบทนิพนธ์เรื่อง General Intelligence , Objectively Determined and Measured โดย สเปียร์แมนจะทดสอบนักเรียนกลุ่มหนึ่งด้วยแบบทดสอบหลายฉบับ และหาความสัมพันธ์ภายใน (Intercorrelation) ระหว่างคะแนนของแบบสอบในแต่ละชุดนั้น และใช้แนวคิดทางจิตวิทยาสร้างรูปแบบของการวิเคราะห์ตัวประกอบแบบ Two - Factor Theory โดยมีสาระว่าความสามารถของมนุษย์จะแบ่งเป็น 2 ตัวประกอบ คือ ตัวประกอบทั่วไป (General Factor) และตัวประกอบเฉพาะ(Specific Factor) ซึ่งจะมีตัวประกอบทั่วไปเพียง 1 ตัว ที่เหลือจะเป็นตัวประกอบเฉพาะ โดยจะเขียนเป็นสมการได้ดังนี้

$$Z_j = a_j G + U_j$$

โดย Z_j แทน ตัวแปรที่วัดได้ (Manifest หรือ Observed Variable)

a_j แทน สัมประสิทธิ์หรือน้ำหนักที่แสดงถึงอัตราความสัมพันธ์ที่ตัวแปร Z_j ที่มีต่อ Latent general - ability G

U_j แทน ตัวประกอบเฉพาะ สำหรับตัวแปร Z_j

แนวคิดและวิธีการวิเคราะห์ตัวประกอบได้รับการพัฒนา โดยนักจิตวิทยาจำนวนหนึ่ง ผู้ซึ่งมีส่วนสำคัญในการวางแนวทางของการวิเคราะห์ตัวประกอบอันได้แก่ เบิร์ต (Cyril Bert)

โฮลซิ่งเจอร์ (Karl Holzinger) เคลลี่(Truman Kelly) เพียร์สัน (Karl Pearson) ทอมสัน (God frey Thamson) และ การ์เนต(Maxwell Garnett) โดยที่บุคคลเหล่านี้ได้พยายามค้นคว้า ต่อจากความพยายามของสเปียร์แมนในเรื่องทฤษฎีสองตัวประกอบ (Two - Factor Theory) ซึ่งวิเคราะห์มาจากความแตกต่างเชิงสี่ (Tetrad Difference) และได้ขยายออกไปจนถึงการพิสูจน์ความคลาดเคลื่อนของการวิเคราะห์

ในราวปี ค.ศ. 1930 เธอร์สโตน(L.L. Thurstone) แห่งมหาวิทยาลัยชิคาโกได้พัฒนาวิธีการวิเคราะห์ตัวประกอบพหุคูณ (Method of multiple factor analysis) ขึ้นมา โดยเชื่อว่ามีตัวประกอบหลายตัว ซึ่งสามารถแทนตัวประกอบ G ของสเปียร์แมนได้ และเริ่มใช้คำว่า "factors" แทนคำว่า " Latent ability " ตั้งแต่นั้นมาการวิเคราะห์ตัวประกอบของเธอร์สโตนได้รับความนิยมมาก และเป็นต้นเหตุสำหรับการพัฒนาวิธีการเซนทรอยด์ (Centroid) และความคิดรวบยอดของโครงสร้างอย่างง่าย (Simple Structure) (Lindeman, Merenda, and Gold, 1980) โดยที่รูปแบบของการวิเคราะห์ตัวประกอบร่วม (Common factor analysis model) มีดังนี้คือ

$$Z_j = a_{j1} F_1 + a_{j2} F_2 + \dots + a_{jn} F_m + d_j U_j$$

โดยที่ ($j = 1, 2, 3, \dots, n$)

n แทน จำนวนตัวแปรที่วัดได้

m แทน จำนวนตัวประกอบร่วม (Common Factors) ซึ่งคาดว่า m น้อยกว่า n

a_{jk} แทน สัมประสิทธิ์ของตัวประกอบโดยทั่วไป ซึ่งจะถูกเรียกว่า น้ำหนักตัวประกอบ (Factor Loading)

โฮลเทลลิ่ง (Harold Hotelling) ได้เสนอรูปแบบสำหรับวิธีวิเคราะห์ตัวแปรอีกแบบหนึ่ง ซึ่งถูกเรียกว่า "Principal Component Analysis" หรือ "Component Analysis Model" ในปี ค.ศ. 1933 โดยจะมีรูปแบบดังนี้ คือ

$$Z_j = a_{j1} F_1 + a_{j2} F_2 + \dots + a_{jn} F_n$$

โดยที่ $j = 1, 2, 3, \dots, n$

เนื่องจากการวิเคราะห์ตัวประกอบร่วม (Common factor Analysis Model) ตามลักษณะของเธอร์สโตนนี้มีผู้นิยมใช้มาก จึงมีผู้คิดค้นเพิ่มเติมอีกหลายคน เช่น การวิเคราะห์แบบภาพพจน์ (Image Analysis) ของกัทแมน (Guttman, 1953) การวิเคราะห์ตัวประกอบแบบคาโนนิคอล (Canonical Factor Analysis) ของราวและแฮร์ริส(Rao, 1955 and Harris, 1962) การวิเคราะห์ตัวประกอบแบบแอลฟา (Alpha Factor Analysis) ของไกเซอร์และเคฟเฟร์ (Kaiser and Caffrey, 1965) และวิธีการน้อยที่สุด (Minres Method) นั้นของฮาร์แมนและโจนส์(Harman and Jones, 1966) ในปี ค.ศ. 1950 เป็นต้นมาเทคนิคทางคอมพิวเตอร์ได้เข้ามามีบทบาทช่วยเหลือในการวิเคราะห์ตัวประกอบในงานวิจัยต่าง ๆ มากขึ้น และในปีค.ศ. 1966 ได้มีเทคนิคทางการวิเคราะห์ตัวประกอบเกิดขึ้นอีกสายหนึ่งนั้นก็คือ Confirmatory Factor Analysis ซึ่งจะมุ่งไปสู่การทดสอบสมมติฐานเกี่ยวกับจำนวนตัวประกอบ โดยคาดว่าผลการวิเคราะห์ที่ได้รับจากกลุ่มตัวอย่าง น่าจะสรุปอ้างอิงไปยังประชากรตัวประกอบได้ และผู้นำในเรื่องนี้คือ บ็อค (Darrel Bock) บาร์คแมน (Bargmann) และ จ็อคโคก (Joreskog)

การวิเคราะห์ตัวประกอบ (Factor Analysis) เป็นเทคนิคทางสถิติที่มีวัตถุประสงค์ในการลดปริมาณของตัวแปรให้มีจำนวนน้อยลง (Variable Reduction) โดยอาศัยโครงสร้างและแบบแผนของความสัมพันธ์ที่มีอยู่ในข้อมูลหรือระหว่างตัวแปรเพื่อให้ง่ายต่อความเข้าใจและทำให้ทราบถึงโครงสร้างและแบบแผนของข้อมูล (Structure and Pattern of Data) นั้น การวิเคราะห์ตัวประกอบจะถูกนำมาใช้ในการตัดสินกำหนดว่าความสัมพันธ์ของตัวแปรนั้นสามารถอธิบายด้วยตัวแปรตามสมมติฐานที่มีจำนวนน้อยลงได้หรือไม่ โดยยึดหลักที่ว่าการที่ตัวแปรหรือข้อมูลต่าง ๆ มีความสัมพันธ์กัน เนื่องจากว่าตัวแปรต่าง ๆ เหล่านี้มีองค์ประกอบร่วมกัน (Common Factors) สามารถสังเกตได้จากการจับกลุ่มกันของตัวแปรหรือค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรนั้นมีความสัมพันธ์กันสูง (Lindeman and et al., 1988)

เทคนิคการวิเคราะห์ตัวประกอบนี้ เป็นประโยชน์มากในการสร้างเครื่องมือทดสอบและการสร้างมาตรวัดตัวแปรประกอบที่ประกอบด้วยตัวแปรต่าง ๆ จำนวนมาก โดยทดสอบว่าตัวแปรต่าง ๆ ที่มารวมกันโดยให้น้ำหนักอย่างถูกต้องหรือไม่ หรือทดสอบว่าตัวแปรต่าง ๆ เหล่านั้นควรจะรวมกันเป็นกลุ่ม ๆ ในลักษณะที่ผู้วิจัยได้กำหนดไว้หรือไม่ การทดสอบทำได้โดยการนำตัวแปรเหล่านั้นมาวิเคราะห์โดยเทคนิคดังกล่าวนี้ แล้วดูแบบแผนการกระจายไปตามองค์ประกอบและน้ำหนักของตัวแปรแต่ละตัวที่มีต่อองค์ประกอบ ถ้าแบบแผนนี้สอดคล้องกับการตัดสินใจของผู้วิจัยในการสร้างตัวแปรประกอบ ก็เป็นการยืนยันว่ามาตรวัดนั้นได้สร้างมาอย่างถูกต้อง

จุดมุ่งหมายของการวิเคราะห์ตัวประกอบ

ในการวิจัยทางสังคมศาสตร์นั้นการใช้เทคนิคการวิเคราะห์องค์ประกอบนี้มีจุดมุ่งหมายที่สำคัญอยู่ 3 ประการ (สุชาติ ประสิทธิ์รัฐสินธุ์ และ ลัดดาวัลย์ รอดมณี, 2527) คือ

1. เพื่อแสวงหาตัวประกอบร่วมที่อธิบายถึงควาสัมพันธ์ร่วมกันระหว่างตัวแปรต่าง ๆ โดยที่จำนวนของตัวประกอบร่วมที่หาได้นั้นจะต้องมีจำนวนน้อยกว่าจำนวนตัวแปรซึ่งจะถูกเรียกว่า Exploratory Factor Analysis

2. เพื่อพิสูจน์ สนับสนุน ตรวจสอบสมมติฐาน เกี่ยวกับโครงสร้างของข้อมูลหรือตัวแปรว่ามีตัวประกอบร่วมกันกี่ตัวประกอบ อะไรบ้าง และตัวแปรแต่ละตัวควรมีน้ำหนักหรืออัตราความสัมพันธ์กับองค์ประกอบมากน้อยเพียงใด ตรงกับที่คาดคะเนไว้หรือไม่ ซึ่งต่างจากวัตถุประสงค์ข้อแรก ซึ่งผู้วิจัยไม่ประสงค์ที่จะทดสอบสมมติฐานเกี่ยวกับการรวมกลุ่มของตัวแปรหรือโครงสร้างของตัวประกอบแต่อย่างใด เพียงแต่ต้องการหาตัวประกอบร่วมจากตัวแปรที่นำมาวิเคราะห์ว่ามีหรือไม่ และมีโครงสร้างเป็นอย่างไร ประกอบด้วยตัวแปรอะไรบ้างเท่านั้น ซึ่งเรียกว่า Confirmatory Factor Analysis

3. เพื่อนำเอาโครงสร้างของความสัมพันธ์ระหว่างตัวประกอบ และตัวแปรเหล่านี้ใช้สร้างคะแนนตัวประกอบ (Factor Score) คะแนนที่ได้นี้เปรียบเสมือนค่าของตัวแปรตัวใหม่ที่ประกอบด้วยตัวแปรเดิมหลาย ๆ ตัวนี้ จะถูกเรียกว่า Composite Variable

ลักษณะของข้อมูลที่ใช้ในการวิเคราะห์ตัวประกอบ

ลักษณะของข้อมูลที่จะนำมาใช้ในการวิเคราะห์ตัวประกอบนั้น จะขึ้นอยู่กับเป้าหมายและรูปแบบของการวิเคราะห์ตัวประกอบ โดยข้อมูลที่ใช้ในการวิเคราะห์จะแบ่งออกเป็น 3 แบบ (สุชาติ ประสิทธิ์รัฐสินธ์ และ ลัดดาวัลย์ รอดมณี, 2527) คือ

1. ข้อมูลที่เป็นตัวแปรบ่งลักษณะของประชากร หรือ ตัวอย่างที่ได้มาจากการสำรวจ เช่น อายุ เพศ การศึกษา รายได้ ซึ่งข้อมูลที่ใช้นั้นจะเป็นตัวแปรที่แสดงค่าต่าง ๆ ของลักษณะของประชากร โดยที่รูปแบบที่ใช้กันส่วนใหญ่ คือ การวิเคราะห์ตัวประกอบที่เรียกว่า ประเภท R (R - type factor analysis)

2. ข้อมูลที่เน้นการวิเคราะห์ความคล้ายคลึง หรือ ความแตกต่างกันของหน่วยซึ่งอาจเป็นบุคคลหรือวัตถุสิ่งของ (Association between individuals or objects) แทนที่จะวิเคราะห์ความสัมพันธ์ หรือความไม่สัมพันธ์กันระหว่างตัวแปร ข้อมูลที่ต้องเตรียม คือ ความสัมพันธ์ระหว่างบุคคลหรือวัตถุสิ่งของ การวิเคราะห์ตัวประกอบประเภทนี้ เรียกว่า Q - type factor analysis ซึ่งยังไม่เป็นที่แพร่หลายในวงการวิจัยทางสังคมศาสตร์

3. เป็นการวิเคราะห์ตัวแปรที่เก็บจากบุคคล หรือวัตถุสิ่งของกลุ่มเดียวกัน 2 ครั้ง และนำเอาคุณสมบัติ หรือตัวแปรมาวิเคราะห์ การวิเคราะห์ตัวประกอบประเภทนี้ เรียกว่า การวิเคราะห์ปัจจัยแบบ 3 ด้าน (Three - model factor analysis) ซึ่งยังไม่เป็นที่แพร่หลายเช่นเดียวกันกับการวิเคราะห์ตัวประกอบแบบ Q

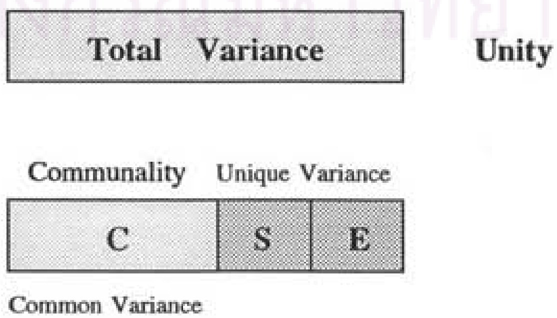
ข้อมูลที่เหมาะสมในการวิเคราะห์ตัวประกอบนั้น จะต้องมีระดับการวัดเป็นมาตราช่วง (Interval Scale) หรือระดับอัตราส่วน (Ratio Scale) แต่ถ้าข้อมูลนั้นเป็นนามบัญญัติ(Nominal Scale) หรือเป็นกลุ่ม(Categorial) จะต้องแปลงข้อมูลนั้นให้มีค่าเป็น 0 หรือ 1 เช่น ผลการสอบรายข้อของนักเรียน ซึ่งตอบถูก = 1 และ ตอบผิด = 0 เป็นต้น

แบบจำลองการวิเคราะห์ตัวประกอบ (Factor Analysis Models)

แบบจำลองการวิเคราะห์องค์ประกอบอาจแบ่งเป็น 2 ประเภทใหญ่ ๆ คือ แบบจำลองที่ว่าด้วยตัวประกอบหลัก (Component Factor Model) และแบบจำลองที่ว่าด้วยตัวประกอบร่วม (Common Factor Model)

1. แบบจำลองที่ว่าด้วยตัวประกอบหลัก (Component factor models) เป็นแบบจำลองที่เน้นเรื่องมิติที่ครอบคลุมความแปรปรวนของตัวแปรทั้งหมด เป็นความพยายามสกัดตัวประกอบจากความแปรปรวนของตัวแปรที่มีอยู่ โดยไม่คำนึงถึงส่วนที่ว่าด้วยความแปรปรวนร่วม หรือความแปรปรวนเฉพาะ และความแปรปรวนคลาดเคลื่อน วิธีการที่ใช้สะกัดตัวประกอบประเภทนี้ได้แก่ Principal components analysis หรือ PC หรือ PA_1

2. แบบจำลองที่ว่าด้วยตัวประกอบร่วม (Common Factor Analysis Model) พัฒนาโดย Spearman (1927) สำหรับตัวประกอบ 2 ตัว ต่อมา Thurstone พัฒนาเป็น Multiple Factor (1935) โดยโมเดลนี้กำหนดให้ตัวแปรแบ่งออกเป็นส่วนย่อย ๆ 2 ส่วนนั้นก็คือ ส่วนที่ร่วมกับตัวแปรอื่น และส่วนเฉพาะของตัวเอง ซึ่งแสดงได้จากภาพดังนี้คือ



ตัวแปร X

แผนภูมิที่ 1 ความแปรปรวนที่ใช้ในการวิเคราะห์ตัวประกอบ

จากแผนภูมิจะเห็นว่าความแปรปรวนทั้งหมด (Total Variance) ของตัวแปรแต่ละตัวนั้น อาจถูกแบ่งออกเป็น ความแปรปรวนที่ร่วมกับตัวแปรอื่น (Common Variance) และความแปรปรวนพิเศษ (Unique Variance) หรือความแปรปรวนที่เหลือ (Residual Variance) ดังเขียนเป็นสมการเส้นตรงได้ดังนี้ คือ

$$\mathbf{T = C + U}$$

โดยที่ **T** แทน ความแปรปรวนทั้งหมด

C แทน ความแปรปรวนที่ร่วมกับตัวแปรอื่น

U แทน ความแปรปรวนพิเศษ หรือ ความแปรปรวนที่เหลือ

สำหรับความแปรปรวนพิเศษ หรือ ความแปรปรวนที่เหลือนั้นยังสามารถแบ่งย่อยเป็น 2 ส่วน คือ ความแปรปรวนเฉพาะตัว (Specific Variance) กับความแปรปรวนที่มาจากความคลาดเคลื่อน (Error Variance) ดังเขียนเป็นสมการเส้นตรงได้ดังนี้ คือ

$$\mathbf{U = S + E}$$

โดยที่ **U** แทน ความแปรปรวนพิเศษ หรือความแปรปรวนที่เหลือ

S แทน ความแปรปรวนเฉพาะตัว

E แทน ความแปรปรวนที่มาจากความคลาดเคลื่อน

ในกรณีวิเคราะห์จากตัวแปรมาตรฐาน (Standard Variable) ความแปรปรวนทั้งหมดของตัวแปรมีค่าเป็น 1.00 ดังนั้นความสัมพันธ์ของความแปรปรวนร่วม ความแปรปรวนพิเศษ ความแปรปรวนเฉพาะ และความคลาดเคลื่อน สามารถเขียนเป็นสมการได้ดังนี้ คือ

$$\mathbf{1 = C + U}$$

หรือ

$$\mathbf{1 = C + S + E}$$

ซึ่งวิธีการที่ใช้สะกัดตัวประกอบประเภทนี้ ได้แก่

2.1 การวิเคราะห์แบบแกนหลัก (Principal axis factoring : PAF หรือ PA_2)

2.2 การวิเคราะห์แบบความเป็นไปได้สูงสุด หรือการวิเคราะห์แบบคโนนิคอล (Canonical factor analysis : ML) ซึ่งพัฒนาโดย ราวและแฮร์ริส (Roa , 1955 and Harris, 1962)

2.3 การวิเคราะห์แบบกำลังสองน้อยที่สุดไม่ปรับน้ำหนัก (Unweight least squares factor analysis : ULS)

2.4 การวิเคราะห์แบบกำลังสองน้อยที่สุดทั่วไป (Generalized least squares factor analysis : GLS)

2.5 การวิเคราะห์แบบแอลฟ่า (Alpha factor analysis) ซึ่งวิธีการนี้ถูกพัฒนาโดย ไกเซอร์และเคฟเฟร์ (Kaiser and Caffrey , 1965)

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

ความแตกต่างระหว่างแบบจำลองทั้งสองในการวิเคราะห์ตัวประกอบ (อุทุมพร จามรมาน , 2532) คือ

1. ในแบบจำลองที่ว่าด้วยตัวประกอบร่วม ผู้วิจัยต้องตั้งข้อตกลงเบื้องต้นเกี่ยวกับค่าความร่วมกัน (Communality : h_j^2) ของตัวแปร แต่ในแบบจำลองที่ว่าด้วยตัวประกอบหลักนั้นไม่จำเป็นต้องทำเช่นนั้น

2. การวิเคราะห์ตัวประกอบซึ่งเริ่มเมตริกซ์สหสัมพันธ์ (Correlation matrix) ค่าในแนวทะแยงของแบบจำลองที่ว่าด้วยตัวประกอบร่วม จะต้องประมาณค่าของค่าความร่วมกันก่อน แต่ในแบบจำลองที่ว่าด้วยตัวประกอบหลัก จะใส่ 1.00 ไว้ในแนวทะแยงของเมตริกซ์สหสัมพันธ์ (R) แทน

3. แบบจำลองทั้งสองจะเหมือนกันในกรณีที่ส่วนเฉพาะ (Specific part) ของตัวแปรมีค่าเป็น 0 นั่นคือ ตัวแปรจะถูกแบ่งออกเป็นส่วนที่ร่วมกัน กับส่วนที่เป็นความคลาดเคลื่อนเท่านั้น

4. แบบจำลองที่ว่าด้วยตัวประกอบร่วม จำนวนตัวแปรจะเท่ากับจำนวนตัวประกอบที่สามารถอธิบายได้ว่า ตัวประกอบจะอธิบายความแปรปรวนของตัวแปร Z_j ได้ทั้งหมด ตัวประกอบในแบบจำลองตัวประกอบร่วมอธิบายความสัมพันธ์ระหว่างตัวแปร ซึ่งจำนวนตัวประกอบจะน้อยกว่าตัวแปรมาก

ขั้นตอนของการวิเคราะห์ตัวประกอบ

การวิเคราะห์ตัวประกอบนี้มีขั้นตอนสำคัญ ๆ ดังต่อไปนี้

1. การกำหนดปัญหาการวิจัย รวบรวมและจัดระเบียบข้อมูล การวิเคราะห์ตัวประกอบนี้จะเหมาะสมกับการวิจัยที่มีวัตถุประสงค์ เพื่อการลดจำนวนข้อมูลหรือการสรุปลักษณะข้อมูล เมื่อได้ปัญหาที่เหมาะสมแล้วก็จะเป็นการกำหนดว่า จะต้องมีตัวแปรอะไรบ้างที่เกี่ยวข้องและจะวัดตัวแปรเหล่านั้นได้อย่างไร ขนาดของกลุ่มตัวอย่างควรเป็นเท่าไร

2. การกำหนดเมตริกซ์สหสัมพันธ์ที่ใช้ในการวิเคราะห์ตัวประกอบว่าต้องการศึกษาถึงสหสัมพันธ์ของตัวแปร หรือที่เรียกว่า R เทคนิค หรือสหสัมพันธ์ระหว่างผู้ตอบ (Respondents) ที่เรียกว่า Q เทคนิค หรืออื่น ๆ นอกจากนี้ยังมีเทคนิค S, T, P และ O ดังแสดงในตารางต่อไปนี้

ตารางที่ 1 การกำหนดเมตริกซ์สหสัมพันธ์ที่ใช้ในการวิเคราะห์ตัวประกอบ

Technique	Factor are loaded by	Indices of association are computed across	Data are collected on
R	Variables	Persons	One occasions
Q	Persons	Variables	One occasions
S	Persons	Occasions	One variables
T	Occasions	Persons	One variables
P	Variables	Occasions	One persons
O	Occasions	Variables	One persons

(Hair , Jr, Anderson & Tatham , 1987)

3. การกำหนด Factor model ว่าจะเป็น Exploratory factor analysis หรือ Confirmatory factor analysis ขึ้นกับจุดมุ่งหมายของการวิจัย

4. การสะกัดตัวประกอบ (Factor extraction)

5. การหมุนแกนตัวประกอบ (Factor rotation)

6. การสร้างสเกลหรือคะแนนตัวประกอบ (Construction of factor score)

แผนภูมิที่ 2 ขั้นตอนในการวิเคราะห์ตัวประกอบในเชิง Explolatory method

ขั้นที่ 1
ปัญหาการวิจัย
มีตัวแปรใดบ้าง
จำนวนตัวแปรเท่าไร
จะวัดตัวแปรได้อย่างไร
ขนาดของกลุ่มตัวอย่าง

↓

ขั้นที่ 2
Correlation matrix
เป็นแบบ R หรือ Q หรือ S หรือ T หรือ P หรือ O

↓

ขั้นที่ 3
Factor model

↓

ขั้นที่ 4
Extraction method
เป็นแบบ PA_1 หรือ PA_2 หรือ Image หรือ Alpha หรือ ML

↓

Unrotated factor matrix
Number of factors

↓

ขั้นที่ 5
Rotated factor matrix
Factor interpretation

↓

ขั้นที่ 6
Factor score
สำหรับการวิเคราะห์ในขั้นต่อไป ได้แก่
Regression
Discriminant analysis
Corrletion

การวิเคราะห์ตัวประกอบหลัก (Principal component analysis)

ภายหลังจาก L.L Thurstone แห่งมหาวิทยาลัยชิคาโก(Chicago) ได้พัฒนา Method of Multiple - Factor Analysis ในปี ค.ศ. 1930 ต่อมาอีก 3 ปี ฮาร์รอด โฮลเทลลิ่ง (Harold Hotelling) ก็ได้เสนอโมเดลสำหรับวิเคราะห์ตัวแปรอีกแบบหนึ่งในปี 1953 เรียก Principal Component Analysis หรือ Component Analysis Model

โฮเทลลิ่ง(Hotelling) ได้เสนอโมเดลของการวิเคราะห์ตัวประกอบหลัก (Principal Component Analysis) ไว้ดังนี้

$$Z_j = a_{j1}F_1 + a_{j2}F_2 + \dots + a_{jn}F_n : (j = 1,2,3, \dots , n)$$

เมื่อ Z_j เป็นObserved variable j (j = 1,2,3, ... , n)

$F_1, F_2, \dots , F_n$ เป็นIndependent Component

a_{ij} เป็นความสัมพันธ์ของตัวแปรที่ j กับ Component ที่ i

จะเห็นได้ว่าจำนวน Component หรือบางคนเรียกว่า Factor จะมีค่าเท่ากับจำนวนตัวแปรที่ถูกสังเกตได้(observe variable) แต่แตกต่างกันที่ตัวแปรที่ถูกสังเกตได้นี้มี Intercorrelation แต่ Component นั้นมีลักษณะเป็น Independent หรือ Orthogonal

การคำนวณเริ่มต้นด้วยเมตริกซ์สหสัมพันธ์ R (Eigenvalue : λ_j) และไอเกนเวคเตอร์ (Eigenvector) ที่สอดคล้องกับค่าไอเกนจะได้ λ_j จำนวน n ตัว และได้ไอเกนเวคเตอร์จำนวน n ชุด เมื่อ normalize eigen vector แล้วสามารถหาค่า

$$a_{j1} = \frac{\alpha_{j1}\sqrt{\lambda_1}}{\sqrt{\alpha^2_{11} + \alpha^2_{21} + \dots + \alpha^2_{n1}}}$$

$\alpha^2_{11} + \alpha^2_{21} + \dots + \alpha^2_{n1}$ เป็น eigen vector coresponding to λ_1 ซึ่ง λ_1 คือ ความแปรปรวนของตัวประกอบที่ j นั้นเอง เนื่องจาก Tr (R) = n ซึ่งเท่ากับจำนวน

ความแปรปรวนทั้งหมด และ $\sum\lambda_j = n$ ฉะนั้น $\lambda_j \times 100$ คือ ค่าร้อยละของความแปรปรวนที่ตัวประกอบ j เป็นตัวอธิบายความสัมพันธ์ได้ ซึ่งจะเห็นว่าการวิเคราะห์ตัวประกอบหลัก (Principal Component Analysis) นี้จะไม่สามารถระบุจำนวนตัวประกอบที่สำคัญได้

ลักษณะของโมเดลของการวิเคราะห์ตัวประกอบหลัก มีดังนี้คือ

1. ใช้เมตริกซ์ความสัมพันธ์ (R) เป็นเมตริกซ์พื้นฐาน
2. ลักษณะสถิติเป็นแบบบรรยาย
3. หลักการจัดกระทำสหสัมพันธ์ระหว่างตัวแปร
4. ไม่มีข้อตกลงเบื้องต้นเกี่ยวกับความแปรปรวนที่ร่วมกับตัวแปรอื่น(Common Variance) และความแปรปรวนพิเศษ (Unique Variance)
5. ไม่คำนึงถึงค่าความร่วมกัน (Communality)
6. จำนวนตัวประกอบเท่ากับจำนวนตัวแปร

Image Factor Analysis (IFA)

Image Analysis พัฒนาโดย หลุยส์ กัทแมน (Louis Guttman) ในปี ค.ศ. 1953 โดยที่การสะกัดตัวประกอบวิธีนี้มีข้อตกลงเบื้องต้นว่า ตัวแปรสามารถแยกออกเป็น 2 ส่วนคือ ส่วนร่วม (Common part) และส่วนพิเศษ (Unique part) ซึ่งจะไม่มีความสัมพันธ์กับตัวแปรอื่น สัดส่วนของทั้งสองส่วนนี้ ส่วนหนึ่งไม่ทราบ (Unknown) และอีกส่วนหนึ่งคาดคะเนบนเมตริกซ์สหสัมพันธ์ (Correlation Matrix) ส่วนร่วมของตัวแปรจะเรียกเงาของตัวแปร (The Image of the Variable) ส่วนพิเศษที่ไม่สามารถคาดประมาณได้ เรียกว่า ต้านเงา (Anti - Image) และจากทฤษฎีของกัทแมน (Guttman)ที่ว่าค่าความร่วมกัน (Communality) ของตัวแปรจะประมาณค่าได้ใกล้เคียงด้วยค่ากำลังสองของสหสัมพันธ์พหุคูณ (The Square of the Multiple Correlation) ระหว่างตัวแปรตัวนั้นกับชุดของตัวแปรที่เหลือนั่นคือการประมาณค่าของค่าความร่วมกัน (Communality) ของตัวแปรใด ๆ จะได้จากเงาของตัวแปรนั้น ซึ่งจะได้รับจากสมการถดถอยของตัวแปรอื่นที่ทำนายตัวแปรนั้น ๆ ดังนั้นคาดคะเนที่ดีที่สุดของส่วนร่วม (Common part) ของตัวแปร จะได้จากเงาของตัวแปรซึ่งจะได้จากการคำนวณ ลักษณะรายละเอียดของ Image Factor Analysis มีดังต่อไปนี้

ทฤษฎีภาพพจน์ (Image Theory)

เมื่อพิจารณาตัวแปรที่สังเกตได้ (Observable) ซึ่งอาจจะเป็นแบบสอบก็ได้ให้ z_j $j = 1, 2, 3, \ldots, n$ แล้วทำให้มันเป็น Standardized คือ มีค่ามัชฌิมเลขคณิต (Mean) เป็น 0 และมีค่าความแปรปรวน (Variance) เป็น 1 ดังนั้นในแต่ละ Test

$$p_j = \sum_{k=1}^{n} w_{jk} z_k \qquad (1)$$

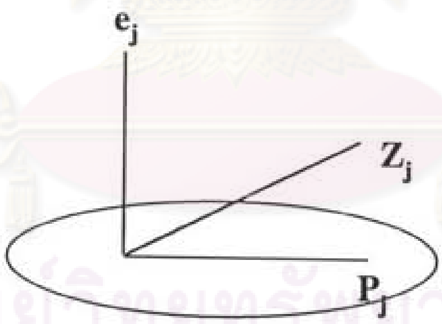
เมื่อ W_{jk} คือ สัมประสิทธิ์ถดถอยที่จะทำนาย Test j จาก Test k ในสมการ Multiple Regression โดยที่ $w_{jj} = 0$

P_j คือ image of test j

เราสามารถหา Anti - Image (e_j) บน test j ได้ซึ่ง

$$e_j = z_j - p_j \qquad (2)$$

ตามหลักเรขาคณิตแล้ว Anti - Image จะตั้งฉากกับ Space of the Remaining tests ซึ่งเสนอได้ดังรูป



SPACE SPANNED BY TEST VECTORS Z_k

โดยที่ $k = 1, 2, \ldots, n$ ซึ่ง $k \neq j$

จากสมการที่ 2 เราสามารถเขียนให้อยู่ในรูปของสมการที่เป็น Fundamental Postulate ได้ดังนี้

$$z = p + e \qquad (3)$$

เมื่อ z คือ เวคเตอร์ของตัวแปร n ตัว ที่เป็นคะแนนมาตรฐาน

p คือ เวคเตอร์ของ image ของตัวแปร n ตัว

e คือ เวคเตอร์ของ anti - image ของตัวแปร n ตัว

จากสมการที่ 1 เราสามารถหา image ของตัวแปรได้จากสมการ

$$p = Wz \qquad (4)$$

เมื่อ W คือ เมตริกซ์ของสหสัมพันธ์ถดถอยพหุคูณ (Matrix of Multiple Regression Coeffecients) ซึ่งแต่ละแถวประกอบด้วยสัมประสิทธิ์สำหรับการทำนายตัวแปรจากตัวแปรที่เหลือ เราสามารถแสดงค่าของ W ได้จากสมการ

$$W = I - S^2 R^{-1} \qquad (5)$$

(Guttman, 1940)

เมื่อ R คือ เมตริกซ์ของสหสัมพันธ์ของ Z's โดยที่ $R = \varepsilon(ZZ')$

S^2 คือ เมตริกซ์แนวทะแยง (diagonal matrix) ของการประมาณความคลาดเคลื่อนของความแปรปรวนหรือความแปรปรวนของ Anti - Image ซึ่งค่า S^2 สามารถหาได้จากสมการ

$$S^2 = [\text{diag } R^{-1}]^{-1} \qquad (6)$$

ถ้าให้ G เป็นความสัมพันธ์ระหว่าง image(Interrelationship of the image) หรือเรียกว่า เมตริกซ์ความแปรปรวนร่วม (Covariance matrix)ของ Image สามารถหาค่า G ได้ดังนี้

$$G = \varepsilon(p\,p') \qquad (7)$$

เมื่อ ε คือ ค่าความคาดหวัง (Expected Value) ดังนี้

$$\begin{aligned} G &= (I - S^2 R^{-1})\, R\, (I - R^{-1} S^2) \\ &= R + S^2 R^{-1} S^2 - 2S^2 \qquad (8) \end{aligned}$$

ในลักษณะเดียวกันถ้าให้ Q เป็นเมตริกซ์ความแปรปรวนร่วมของ Anti - Image เราสามารถหาค่าของ Q ได้ดังนี้

$$
\begin{aligned}
Q &= \varepsilon\,(e\,e') \\
&= \varepsilon\,(S^2R^{-1}Z\,Z'R^{-1}S^2) \\
&= S^2R^{-1}S^2 \qquad (9)
\end{aligned}
$$

จากสมการที่ 6 และ 9 จะได้

$$\text{diag}\ Q = S^2 \qquad (10)$$

หรือกล่าวว่า S^2 คือ เมตริกซ์แนวทะแยงของความแปรปรวนของ Anti - Image

จากสมการที่ 8 และ 9 สามารถทำให้เป็น Fundamental theorem ของการวิเคราะห์ภาพพจน์(Image Analysis) ได้ดังนี้

$$R = G - Q + 2S^2 \qquad (11)$$

นั่นคือเมตริกซ์ความสัมพันธ์ของตัวแปร (R) มีค่าเท่ากับเมตริกซ์ความแปรปรวนร่วมของ Image (G) ลบด้วย เมตริกซ์ความแปรปรวนร่วมของ Anti - Image

จากที่เราเรียก Image และ Anti - Image นั่นเป็นเพียง Partial image และ Partial anti - image เท่านั้น เมื่อพิจารณาถึง total image และ total anti - image เราสามารถที่จะหาได้ เมื่อเรานำตัวแปรในเรื่องนั้นมาวิเคราะห์ได้ทั้งหมด total image ของตัวแปร (π_j) มีค่าดังนี้

$$\pi_j = \lim_{n \to \infty} \sum_{k=1}^{n} w_{jk}\, z_k \quad \text{เมื่อ} \quad w_{jj} = 0 \qquad (12)$$

เมื่อ w_{jk} คือ สัมประสิทธิ์ถดถอยในเอกภพสัมพัทธ์ของตัวแปร (Universe regression Coefficient)

ค่าลิมิตในสมการที่ 12 เป็นค่าลิมิตทางคณิตศาสตร์ (Mathematical Limit) ซึ่งกัทแมน (Guttman) ได้พิสูจน์ว่าสามารถหาค่าได้เสมอ ในขณะเดียวกันเราสามารถหา total anti-image ได้ดังนี้

$$\varepsilon_j = Z_j - \pi_j \qquad (13)$$

การนำมาประยุกต์ใช้กับการวิเคราะห์ตัวประกอบ (Application to Factor Analysis)

ให้ r เป็นจำนวน Common factors จาก n ตัวแปร สามารถแสดงได้ว่า

$$\lim_{n \to \infty} \frac{r}{n} = 0 \qquad (14)$$

ดังนั้น Image Analysis ในเอกภพสัมพัทธ์ (total image analysis) และ factor analysis จึงเป็นสิ่งเดียวกัน ซึ่งแสดงความสัมพันธ์ได้ดังนี้

$$P_j \rightarrow \pi_j = c_j \qquad (15)$$

เมื่อ C_j คือ factor - analytic common - part ของ Z_j ในทำนองเดียวกัน

$$e_j \rightarrow \varepsilon_j = y_j \qquad (16)$$

เมื่อ y_j คือ Unique - part ของ z_j ($z_j = c_j + y_j$)

จากสมการที่ 15 และ 16 มีข้อตกลงเบื้องต้น(Assumption) ที่ว่า ความแปรปรวนร่วม (Covariance) มีค่าเท่ากับ 0

$$q_{jk} = Cov(e_j, e_k) \quad Cov(\varepsilon_j, \varepsilon_k) = Cov(y_j, y_k) = 0 \qquad (17)$$

$$Cov(p_j, e_k) \rightarrow Cov(\pi_j, \varepsilon_k) = Cov(c_j, y_k) = 0 \qquad (18)$$

และตามทฤษฎีของ Factor Analysis

$$q_{jj} = Var(p_j) \rightarrow Var(\pi_j) = Var(c_j) = h^2_j \qquad (19)$$

$$q_{jj} = Var(e_j) \rightarrow Var(\varepsilon_j) = Var(y_j) = u^2_j \qquad (20)$$

$$q_{jk} = Cov(p_j, p_k) \rightarrow Cov(\pi_j, \pi_k) = Cov(c_j, c_k) = r_{jk} \qquad (21)$$

เมื่อ h^2_j คือ ค่าความร่วมกัน (Communality) ของตัวแปร j ($j \neq k$)

u^2_j คือ Uniqueness ของ Test j ซึ่ง $h^2_j + u^2_j = 1$

และ r_{jk} คือ สหสัมพันธ์ความสัมพันธ์ระหว่าง Test j กับ k

จากสมการที่ 14 จะได้

$$G \rightarrow R - U^2 \qquad (22)$$

$$\text{diag } Q = S^2 - U^2 \qquad (23)$$

$$Q \rightarrow U^2 \qquad (24)$$

เมื่อ U^2 เป็นเมตริกซ์แนวทะแยง ของ Uniquenesses

G เป็นตัวประมาณค่า $R - U^2$ ที่ดี เรียกว่า reduced Correlation matrix (คือเมตริกซ์ความแปรปรวนร่วมของ Common part ของตัวแปรสังเกตได้จากสมาชิกนอกแนวทะแยงของเมตริกซ์ (Off - diagonal elements) ในทางตรงกันข้าม ถ้า Q ไม่เป็นเมตริกซ์แนวทะแยง G จะเป็นตัวประมาณค่าที่ไม่ดี ในบางครั้งการวิเคราะห์ตัวประกอบอาจจะไม่เหมาะกับข้อมูลที่เราศึกษา โดยที่เขียนจากข้อมูลที่ไม่คลุมทั้งเอกภพสัมพันธ์ของตัวแปรจึงทำให้ image ไม่ใช่ตัวประกอบที่ดีพอ สำหรับการวิเคราะห์ตัวประกอบแต่เรายังยอมรับว่าเป็นตัวประมาณค่าที่ดีที่สุดของค่าความร่วมกัน(Communality)

ปัญหาต่อมาก็คือเรื่องของจำนวนตัวประกอบ(factor) ซึ่ง แฮร์ริส(Harris, 1962) ได้ให้คำตอบเรื่องนี้ โดยศึกษาจากเรื่อง Canonical factor Analysis ของราว(Rao) 1955 ซึ่งเป็นวิธีการทางสถิติ คำตอบของราวน์(Rao) คล้ายคลึงกับวิธีการ Maximum - likelihood (ML) ของ ลอร์เร่ (Lawley) 1940 ราวน์ได้ให้คำตอบของ Eigenequation ไว้ดังนี้

$$[U^{-1} R U^{-1} - \beta I]\,\varepsilon = 0 \qquad (25)$$

ถ้าเราให้ S^2 เป็น ค่าประมาณค่าของเมตริกซ์แนวทะแยงของ Uniqueness (U^2) จากสมการที่ 25 จะได้สมการใหม่ดังนี้

$$[S^{-1} R S^{-1} - bI]\,x = 0 \qquad (26)$$

เมื่อ b คือ ค่าประมาณของ β

x คือ ค่าประมาณของ ε

คำตอบของสมการที่ 26 แสดงด้วย $XB^{1/2}$ ซึ่งเป็นแกนหลักของการวิเคราะห์ตัวประกอบของ

$$R* = S^{-1} R S^{-1}$$

เมื่อ X เป็นเมตริกซ์ของ Unit - lenght eigenvectors ของ R*

B คือ เมตริกซ์แนวทะแยงของ ค่าไอเกน (Eigenvalues) ของ R* จะได้ factor matrix F_r ของ R ดังนี้

$$F_r = S X B^{1/2} \qquad (27)$$

จาก Image Analysis แฮร์ริสได้พบว่า ไอเกนเวคเตอร์ (eigenvectors) ของ $G* = S^{-1} G S^{-1}$ เหมือนกับไอเกนเวคเตอร์ (eigenvectors) ของ R* แต่ค่าไอเกน(Eigenvalues) ของ G* คือ $[(B - I)^2 B^{-1}]$ ดังนั้น Factoring F_g ของ G คือ

$$Fg = SX [(B - I)^2 B^{-1}]^{1/2} \qquad (28)$$

จากสมการที่ 27 และ 28 จะได้ $F_g = F_r D$

เมื่อ D เป็นเมตริกซ์แนวทะแยง

$$D = I - B^{-1} \qquad (29)$$

โมเดลการวิเคราะห์แบบภาพพจน์ (Image Factor Analysis) นี้จะใช้ค่าสหสัมพันธ์พหุคูณยกกำลังสองระหว่างตัวแปรนั้นกับตัวแปรที่เหลือเป็นค่าความร่วมกัน (Communality : h_j^2) โดยนำค่าความร่วมกันไปแทนค่าสมาชิกในแนวทะแยงของเมตริกซ์ แล้วนำไปหาค่าไอเกน จะได้จำนวนตัวประกอบที่สำคัญที่มีค่าไอเกนมากกว่า 1 ตามแนวของ Henry F.Kaiser ที่ให้ไว้

การหมุนแกนตัวประกอบ (Factor rotation)

หลังจากสะกัดตัวประกอบ มักพบว่าตัวประกอบแรกที่หาได้จะอธิบายความแปรปรวนร่วมระหว่างตัวแปรได้มากกว่าตัวประกอบตัวต่อมาตามลำดับ ตัวประกอบที่สองจะอธิบายความแปรปรวนที่เหลือจากการอธิบายด้วยตัวประกอบแรก ตัวประกอบที่สามจะอธิบายความแปรปรวนที่เหลือที่เหลือจากการอธิบายด้วยตัวประกอบ 2 ตัวแรก เช่นนี้เรื่อยไป จากผลการสะกัดตัวประกอบ ในบางครั้งพบความสลับซับซ้อนของตัวประกอบ ในกรณีที่ตัวแปรตัวหนึ่งมีน้ำหนักตัวประกอบ (Factor loading) บนตัวประกอบมากกว่า 1 ตัว ยิ่งมีความสลับซับซ้อนของตัวประกอบมากเท่าใด ความยุ่งยากในการแปลความหมายของตัวประกอบและตัวแปรก็มีมากเท่านั้น ดังนั้น การเปลี่ยนแปลงตำแหน่งของแกน หรือ มิติ หรือ ตัวประกอบบ้างเล็กน้อยอาจได้ภาพที่ดีขึ้น การหมุนแกนจึงเป็นการเปลี่ยนตำแหน่งของข้อมูลตัวแปรให้สัมพันธ์กับตัวประกอบในลักษณะที่ชัดเจนขึ้น เพื่อลดความสลับซับซ้อนของตัวแปรให้ต่ำลงและทำให้ตัวแปรแต่ละตัวจะได้มีน้ำหนักบนตัวประกอบเพียงตัวเดียวเท่านั้น

การเปลี่ยนตำแหน่งของข้อมูลตัวแปรให้สัมพันธ์กับตัวประกอบทำได้ 2 แบบ คือ

1. ยังคงให้แกนตัวประกอบตั้งฉากซึ่งกันและกันอยู่ แบบนี้เรียกว่า เป็นการหมุนแกนแบบออโธกอนอล (Orthogonal)

2. แกนตัวประกอบไม่ต้องตั้งฉากกัน (แต่ไม่ควรทับกัน) แบบนี้เรียกว่า การหมุนแกนแบบออบลีค (Oblique)

การที่ตัวประกอบทำมุม 90° ซึ่งกันและกัน แสดงความเป็นอิสระทางสถิติของตัวประกอบเหล่านั้น ในกรณีที่ตัวประกอบทำมุมมากกว่าหรือน้อยกว่า 90° แสดงให้เห็นว่า ตัวประกอบทั้งสองเริ่มสัมพันธ์กัน

ลักษณะการหมุนแกนแบบออโธกอนอลกับออบลีค (อุทุมพร จามรมาน, 2532)

แบบออโธกอนอล (Orthogonal)

1. ผลคูณภายในของน้ำหนักตัวประกอบเป็น 0

2. คะแนนตัวประกอบเป็นอิสระเชิงเส้นตรงและไม่สัมพันธ์กัน นั่นคือเมตริกซ์สหสัมพันธ์ของตัวประกอบ คือ ไอเดนติตี้เมตริกซ์ (Identity matrix)

3. ลำดับที่ของตัวประกอบที่หมุนแกนแล้ว อาจแตกต่างจากที่ยังไม่ได้หมุนแกน

4. ผลคูณภายในของเมตริกซ์ตัวประกอบที่หมุนแกนแล้ว มีค่าเท่ากับ ผลคูณภายในของเมตริกซ์ตัวประกอบที่ยังไม่ได้หมุนแกน

แบบออบลีค (Oblique)

1. คะแนนตัวประกอบมีความสัมพันธ์ซึ่งกันและกัน

2. มีเมตริกซ์ใหม่ที่แยกออกจากกัน คือ เมตริกซ์โครงสร้างตัวประกอบ กับเมตริกซ์แบบแผนตัวประกอบ

3. น้ำหนักตัวประกอบมีความหมาย คือ สหสัมพันธ์ระหว่างตัวแปร กับตัวประกอบในเมตริกซ์แบบแผนของแกนอ้างอิง

4. น้ำหนักตัวประกอบอาจมีค่ามากกว่า 1.0

5. ค่าความร่วมกันไม่สามารถคำนวณจากน้ำหนักตัวประกอบได้โดยตรง

6. ไม่สามารถคำนวณความแปรปรวน อันเนื่องจากตัวประกอบได้โดยตรง

เกณฑ์การหมุนแกน

ในการหมุนแกนแบบออโธกอนอล หรือแบบออบลีคก็ตาม ก็ควรมีเกณฑ์การหมุนแกน เกณฑ์การหมุนแกนมีหลายชนิด เช่น

1. เกณฑ์โครงสร้างอย่างง่าย (Simple structure)

เกณฑ์ที่นิยมและยอมรับอย่างกว้างขวางในการวิเคราะห์ตัวประกอบ และทำให้การแปลความหมายตัวประกอบได้อย่างชัดเจนขึ้น คือเกณฑ์โครงสร้างอย่างง่าย โดยการหมุนแกนเพื่อกระจายความแปรปรวนจากตัวประกอบแรก ไปยังตัวประกอบตัวสุดท้าย โดยเฉพาะตัวแปรที่สัมพันธ์กัน ก็ให้อยู่บนตัวประกอบเดียวกัน และที่ไม่สัมพันธ์กันก็ไม่ควรอยู่บนตัวประกอบนั้น เกณฑ์การหมุนแกนโครงสร้างอย่างง่ายนี้พัฒนาโดย เธอร์สโตน (Thurstone , 1947) และอีวีริท และดัน (Everitt and Dunn ,1991) มีหลักดังนี้

1.1 ในแต่ละแถวของ Pattern matrix A ควรจะมีอย่างน้อยที่สุด 1 ตัวที่มีค่าเป็นศูนย์

1.2 ถ้ามีตัวประกอบร่วม (Common factors) m ตัว ในแต่ละคอลัมน์ของ Pattern matrix A ควรมีน้ำหนักที่เป็นศูนย์อย่างน้อย m ตัว

1.3 สำหรับทุกคู่ของคอลัมน์ของ Pattern matrix A ควรจะมีหลาย ๆ ตัวแปรที่มีน้ำหนักเป็นศูนย์ในคอลัมน์หนึ่งและไม่เป็นศูนย์ในอีกคอลัมน์หนึ่ง

1.4 สำหรับทุกคู่ของคอลัมน์ใน Pattern matrix A ควรมีสมาชิกที่มีน้ำหนักเป็นศูนย์ทั้งสองคอลัมน์จำนวนมาก ๆ ถ้ามีตัวประกอบร่วม (Common factor) เป็น 4 ตัวประกอบหรือมากกว่า

1.5 สำหรับทุกๆคู่ของคอลัมน์ใน Pattern matrix A ควรมีจำนวนตัวแปรจำนวนน้อย ๆ ที่มีสมาชิกไม่เท่ากับศูนย์ ปรากฏอยู่บนทั้งสองคอลัมน์

2. เกณฑ์หมุนแกนเพื่อให้ตัวประกอบผ่านกลุ่มตัวแปร

การหมุนแกนแบบโครงสร้างอย่างง่าย ในบางครั้งมิได้เป็นการนำตัวประกอบเข้าใกล้กลุ่มของเวคเตอร์ตัวแปร ศาสตราจารย์ไทรออน ได้สนับสนุนวิธีการหมุนแกนเพื่อให้ตัวประกอบผ่านกลุ่มตัวแปรนี้ว่า เป็นเกณฑ์การหมุนที่ดี

3. เกณฑ์หมุนแกนเพื่อให้สอดคล้องกับผลในอดีต

การหมุนแกนตัวประกอบให้สอดคล้องกับผลในอดีต Castell ได้ใช้การวิเคราะห์ตัวประกอบในเรื่องสติปัญญา โดยเขาศึกษาจากประชากรทั่วไปกลุ่มหนึ่งและกับกลุ่มนิสิตนักศึกษาอีกกลุ่มหนึ่ง ถึงแม้ว่าทั้งสองกลุ่มจะมีระดับสติปัญญาต่างกัน แต่แบบแผนตัวประกอบย่อมคล้ายกับการวิเคราะห์ตัวประกอบของกลุ่มแรก จึงเป็นแนวทางในการหมุนแกนแก่กลุ่มตัวอย่างที่สอง

4. เกณฑ์หมุนแกนเพื่อให้ผลตรงกับสมมติฐาน

คล้ายกับเกณฑ์การหมุนแกนเพื่อให้ตัวประกอบผ่านกลุ่มตัวแปร แต่ต่างกันที่สมมติฐานนี้ ผู้วิจัยอาจนำมาจากทฤษฎีก็ได้ เป็นการพยายามที่จะหมุนแกนตัวประกอบให้ตรงกับลักษณะธรรมชาติที่คาดการณ์ไว้ล่วงหน้า

5. เกณฑ์การหมุนเพื่อให้ดูง่ายขึ้น

การได้จำนวนตัวประกอบน้อยลง เมื่อเปรียบเทียบกับจำนวนข้อมูลมาก ๆ เป็นการช่วยให้ตีความหมายง่ายขึ้น

6. เกณฑ์การหมุนเพื่อให้ความแปรเปลี่ยนคงที่ (Invariance)

จุดมุ่งหมายอีกอย่างหนึ่งของการวิเคราะห์ตัวประกอบ คือ การสามารถอ้างอิงผลของตัวประกอบได้มากที่สุด โดยเฉพาะคำตอบน่าจะมีเพียงคำเดียว การจัดกระทำให้ตัวแปรที่ควรจะอยู่ที่ตัวประกอบลงให้เหลือเพียงแบบเดียว (Factor invariance)

7. เกณฑ์ค่าบวกจำนวนมาก

แม้ว่าสหสัมพันธ์ในเมตริกซ์จะเป็นค่าบวกและศูนย์(ไม่มีค่าลบเลย) ตัวประกอบตั้งแต่ตัวที่สอง จะมีน้ำหนักเป็นค่าบวกและลบผสมกัน ตามข้อกำหนดของวิธีเซนทรอยด์ที่ว่าผลบวกของน้ำหนักของตัวประกอบตัวต่อไปจะเป็นศูนย์ เนื่องจากสหสัมพันธ์เป็นค่าบวกหมดตัวประกอบที่หมุนแกนแล้วจะไม่ค่าลบเลย

วิธีการหมุนแกนแบบออโธกอนอล มีเทคนิคที่สำคัญ คือ

1. วิธีควอติแมกซ์ (Quartimax)

หลักของเทคนิคนี้คือ การลดกำลังสี่ของน้ำหนักตัวประกอบให้น้อยที่สุด จึงมีวิธีเสนอวิธีวิเคราะห์ตัวประกอบโดยใช้หลักนี้หลายคน เช่น เฟอร์กูสัน (Ferguson , 1954) นีซัสและวิลลี่ (Neusaus and Wrigley , 1954) แคลรอล (Carroll , 1953) แซนเดอร์ (Saunders, 1960) l

หลักของวิธีนี้ก็คือ การหมุนแกนทีละแกนเข้าหาจุด การมีจำนวนจุดหลายจุดเป็นการลดผลคูณของน้ำหนักตัวประกอบทีละคู่ที่เป็นไปได้สำหรับตัวแปรแต่ละตัว รวมค่าเหล่านี้สำหรับทุกตัวแปร เพื่อให้ได้ฟังก์ชั่นของเมตริกซ์ตัวประกอบ

ฮาร์แมน (Harman , 1967) ชี้ให้เห็นว่า การลดค่าผลรวมของผลคูณเช่นนี้ อาจเขียนเป็นสมการได้คือ

$$Q = \sum_{i}^{n} \sum_{j}^{m} a_{ij}^{4} \qquad (30)$$

เมื่อ Q คือ ค่าที่ลดลงต่ำสุด

a_{ij} คือ น้ำหนักตัวประกอบของตัวแปร x_i บนตัวประกอบ I การลดค่าให้ต่ำสุดของสมการนี้ทำให้ค่าที่สูงเพิ่มสูงขึ้น ส่วนค่าที่ปานกลางอยู่แล้วก็ลดลง ผลที่ได้รับให้ภาพพจน์โครงสร้างอย่างง่าย และตัวประกอบตัวแรกได้ค่าความแปรปรวนสูงกว่าการหมุนด้วยมือ

2. การหมุนแกนแบบวาริแมกซ์ (Varimax)

การหมุนแกนมุมฉากแบบวาริแมกซ์ เป็นวิธีการที่เสนอโดย ไกเซอร์ (Kaiser, 1956) ซึ่งวิธีการนี้ให้ความสำคัญต่อปัจจัย (หรือคอลัมน์)แต่ละปัจจัย แทนที่จะหาค่าสูงสุดของการผันแปรของน้ำหนักของปัจจัยของตัวแปรตามแถว การหมุนแกนแบบวาริแมกซ์มุ่งไปที่ความแตกต่างหรือความผันแปรของแต่ละปัจจัย โดยพยายามทำให้ปัจจัยแต่ละคอลัมน์แตกต่างกันให้มากที่สุด ค่าที่ต้องการพยายามให้ได้สูงสุดคือ

$$V_j^{*} = \frac{r \sum_{j=1}^{n} (a_{ij}^2)^2 - (\sum_{j=1}^{n} a_{ij}^2)^2}{n^2} \qquad (31)$$

เมื่อ a_{ij} คือ น้ำหนักของตัวแปร i บนปัจจัย j

n คือ จำนวนตัวแปร

V_j^{*} คือ ค่าของการผันแปรของปัจจัย j

ในที่นี้ a_{ij} ไม่ใช่ค่าคงที่ของอัตราการร่วมกัน เพราะเป็นน้ำหนักเชิงปัจจัยของตัวแปรบนแต่ละปัจจัยให้ค่า V_j^{*} ของทุกปัจจัยเท่ากับ V ตามสมการข้างล่าง

$$V^{*} = \sum_{j=1}^{r} V_j^{*} = \sum_{j=1}^{r} \frac{n \sum_{i=1}^{n} (a_{ij}^2)^2 - (\sum_{i=1}^{n} a_{ij}^2)^2}{n^2} \qquad (32)$$

เพื่อมิให้ a_{ij} ของตัวแปรตัวใดมีผลลัพธ์มากเกินไป หารน้ำหนัก a_{ij} ด้วยค่าความร่วมกัน h_j^2 ของตัวแปรตัวนั้น ดังนั้นควรใช้ a_{ij}^2 / h_j^2 แทน a_{ij} ในสมการที่ (1) และ a_{ij}^4 / h_j^4 แทน a_{ij}^4

จากสมการที่ (32) จะเห็นว่าเมื่อค่าของ V สูงก็ต่อเมื่อความแตกต่างระหว่างปัจจัยมีมาก ทำให้สะดวกแก่การอธิบายปัจจัยแต่ละปัจจัยได้ดีกว่าในกรณีที่ปัจจัยมีความคล้ายคลึงกัน

3. การหมุนแกนแบบอิควาแมกซ์ (Equamax)

เป็นวิธีผสมผสานระหว่างวิธีวาริแมกซ์กับวิธีควอติแมกซ์ เป็นการลดทั้งจำนวนตัวแปรและจำนวนตัวประกอบ

วิธีการหมุนแกนแบบออบลีค (Oblique rotation)

การหมุนแกนแบบออบลีค เป็นการหมุนแกนที่ไม่ได้ตั้งฉากกัน นั่นคือ แกนทั้งสองทำมุมมากกว่า 90° การหมุนแกนแบบออบลีคเกี่ยวข้องกับเมตริกซ์หลายตัว คือ

- เมตริกซ์สหสัมพันธ์ระหว่างตัวประกอบ (ϕ)

- เมตริกซ์โครงสร้าง (Structure matrix) หรือ S คือเมตริกซ์ที่แสดงค่าสหสัมพันธ์ระหว่างข้อมูลตัวแปรกับตัวประกอบออบลีค

- เมตริกซ์แบบแผน (Pattern matrix) หรือ P คือเมตริกซ์ที่แสดงค่าน้ำหนักตัวประกอบของข้อมูลตัวแปรเป็นตัวประกอบร่วมออบลีค

การหมุนแกนแบบออบลีค มีเทคนิคที่สำคัญคือ

1. การหมุนแกนแบบออบลิแมกซ์ (Oblimax)

แซนเดอร์ (Saunders , 1961)ได้พัฒนาวิธีการหมุนแกนแบบออบลิแมกซ์ จากแกนออโธกอนอลเป็นแกนออบลีค เขาพยายามหาวิธีหมุนแกนตัวประกอบ เพื่อให้น้ำหนักตัวประกอบมีค่าไม่สูงก็ต่ำ มิใช่ค่าปานกลาง โดยใช้ฟังก์ชั่นดังนี้

$$K = \sum_{i=1}^{n}\sum_{j=1}^{m} a_{ij}^4 \;/\; \left(\sum_{i=1}^{n}\sum_{j=1}^{m} a_{ij}^2\right)^2$$

เมื่อ a_{ij} คือ น้ำหนักแกนอ้างอิงออบลีค

วิธีออบลิแมกซ์ให้ภาพพจน์ใกล้เคียงกับโครงสร้างอย่างง่าย ถ้าข้อมูลจัดอยู่ในภาพพจน์โครงสร้างอย่างง่าย

2. การหมุนแกนแบบควอติมิน (Quartimin)

แครอล (Carroll , 1953) เพื่อลดค่าผลรวมของผลคูณภายในของน้ำหนักแกนอ้างอิง สมการคือ

$$Q = \sum_{j<k=1}^{m} \sum_{i=1}^{n} a_{ij}^2\, a_{ik}^2$$

เมื่อ a_{ij} คือ น้ำหนักแกนอ้างอิงออบลีค j และ k คือ ตัวประกอบ

ในการคำนวณการลดค่า **Q** เกี่ยวกับคำตอบจากค่าไอเกนสำหรับแต่ละตัวประกอบโดยคุมตัวประกอบอื่นให้คงที่ ทำซ้ำเช่นนี้สำหรับตัวประกอบแต่ละตัวจนครบทุกตัว และเริ่มทำซ้ำเพื่อให้ลดค่าจนน้อยที่สุด

3. การหมุนแกนแบบโควาริมิน (Covarimin)

ไกเซอร์ (Kaiser, 1958) ได้เสนอวิธีหมุนแกนออบลีคต่อจากเกณฑ์วาริแมกซ์ โดยทำฟังก์ชั่นต่อไปนี้ให้มีค่าน้อยที่สุด

$$C = \sum_{j<k=1}^{m} \left[n \sum_{i=1}^{n} (a_{ij}^2 / h_i^2)(a_{ij}^2 / h_i^2) - \{\sum_{i=1}^{n} a_{ij}^2 / h_i^2 \}\{ \sum_{i=1}^{n} a_{ij}^2 / h_i^2 \} \right]$$

เมื่อ a_{ij} คือ น้ำหนักแกนออบลีคอ้างอิง i

j คือ ตัวประกอบ

ผลจากควาริมินมักจะให้ค่าความสัมพันธ์ของตัวประกอบมีค่าต่ำ และนั่นคือทำให้ได้ค่าใกล้เคียงกับคำตอบจากวาริแมกซ์

4. ไบควอติมิน (Biquaretimin)

แครอล (Carrol , 1957) อาศัยวิธีของควอติมิน และโควาริมินมาพิจารณาคือ

$$B = \frac{Q + C}{n}$$

เมื่อ **Q** คือ Quartimin function

C คือ Covarimin function

n คือ จำนวนตัวแปร

6. แมกซ์เพลน (Maxplane)

คาทเทล และมัลเรล (Cattell and Murel , 1960) เสนอวิธีการที่แตกต่างจากสองวิธีแรก โดยพิจารณาจาก Hyperplane และพิจารณาตัวประกอบทีละคู่ โดยนับจุดของตัวแปรที่ตกอยู่ใน Plane ที่มีค่าความ + 1.00 แล้วนับจำนวน Hyperplane ที่มีจุดเหล่านี้ตกอยู่ภายใต้ช่วง 2 S.D.

7. โปรแมกซ์ (Promax)

เฮนดิสันและไว้ซ์ (Hendrickson and White , 1964) เสนอวิธีหมุนแกนที่ทำได้อย่างรวดเร็ว โดยการทำค่าน้ำหนักตัวประกอบทั้งแถวและคอลัมน์ให้เป็นค่าปรกติตามภาพที่ต้องการ

8. หมุนแกนตามเป้าหมาย (Target rotation)

เป็นการหมุนแกนออบลีคตามภาพพจน์ที่กำหนดไว้ล่วงหน้าตามข้อมูล ให้ค่าไม่เปลี่ยนแปลง (Invariance)

ตอนที่ 3 สัมประสิทธิ์สหสัมพันธ์ที่ใช้ในงานวิจัย

ในปัจจุบันนี้คุณภาพของการวิจัย โดยการหาค่าสหสัมพันธ์นี้มิได้ขึ้นอยู่กับการออกแบบที่ซับซ้อน หรือกลวิธีการสมัยใหม่ในการหาค่าสัมประสิทธิ์สหสัมพันธ์แต่จะขึ้นอยู่กับการวางแผน และทฤษฎีที่สร้างขึ้นเพื่อกำหนดไว้เป็นสมมติฐาน ในสมัยก่อนการวิจัยทางการศึกษาจำนวนมากที่ใช้วิธีหาสหสัมพันธ์ โดยหาข้อมูลเกี่ยวกับคะแนนของกลุ่มหนึ่งมามาก หรือน้อยตามต้องการ แล้วนำมาหาค่าสัมประสิทธิ์สหสัมพันธ์เพื่อหาความสัมพันธ์กันนั้นการศึกษานี้ไม่จัดเป็นการวิจัยเชิงวิทยาศาสตร์ (Borge, 1965) ทั้งนี้เนื่องจากว่าไม่มีพื้นฐานหลักทางทฤษฎีมายืนยัน และไม่มีการทดสอบสมมติฐาน ซึ่งอาจจะถือเป็นเพียงการลองผิดลองถูกเท่านั้น อันที่จริงค่าสัมประสิทธิ์สหสัมพันธ์นี้จะบอกให้ทราบว่ามีความสัมพันธ์กันหรือไม่มากน้อยเพียงใด แต่จะไม่แสดงให้ทราบถึงเหตุและผลแห่งความสัมพันธ์กัน โดยไม่เกี่ยวข้องกับสาเหตุหรือผลที่จะบังเกิดขึ้น ไม่เกี่ยวข้องกับเหตุผล หรือ ความสัมพันธ์ในด้านส่วนบุคคลของสิ่งใดสิ่งหนึ่งเพียงสิ่งเดียวในชุดนั้น และจะไม่สามารถแสดงให้ทราบถึงขนาดใหญ่เล็กของข้อมูลแต่ละชุด แต่ค่าสัมประสิทธิ์สหสัมพันธ์จะแสดงให้เห็นเพียงความสัมพันธ์ของส่วนรวมทั้งชุด หรือกลุ่มต่อกลุ่มเป็นเกณฑ์ นอกจากนี้ยังแสดงให้ทราบถึงสัดส่วนความสัมพันธ์กันของข้อมูลทั้งสองชุด ว่ามีลักษณะการกระจายคะแนนเรียงลำดับกันเป็นระเบียบทำนองเดียวกันหรือไม่ อีกทั้งยังใช้ทำนายผลหรือแนวโน้มที่เกิดขึ้น กล่าวคือ เป็นการประมาณค่าของข้อมูลชุดหนึ่งที่จะเกิดขึ้นจากค่าต่าง ๆ ของข้อมูลอีกชุดหนึ่งที่ทราบค่าแล้ว ซึ่งลินเดอร์แมน มีเรนดา และโกลด์ (Lindeman, Merenda and Gold, 1980) ได้กล่าวไว้ว่า สหสัมพันธ์ (Correlation) คือค่าวัดความมากน้อยในความสัมพันธ์เชิงเส้นตรงระหว่างตัวแปรสุ่มสองตัวแปร โดยที่ช่วงของค่าสัมประสิทธิ์สหสัมพันธ์จะมีค่าตั้งแต่ -1 จนถึง 1 และค่าสัมประสิทธิ์สหสัมพันธ์นั้นมีค่าเท่ากับ 1 ย่อมแสดงว่า ความสัมพันธ์ของตัวแปรคู่นี้จะมีลักษณะที่สมบูรณ์แบบ (Perfect relation) ซึ่งทั้งสองกรณีนี้มักจะไม่ค่อยพบในการคำนวณจากข้อมูลจริง แต่ค่าสัมประสิทธิ์สหสัมพันธ์นี้อาจจะมีค่าบวกหรือลบก็ได้แล้วแต่ความสัมพันธ์ของตัวแปรคู่ที่นำมาศึกษาว่าอาจจะเป็นในลักษณะที่ตามกัน (Positive relation) ซึ่งความสัมพันธ์ของตัวแปรคู่นั้น อาจจะเพิ่มขึ้นหรือลงที่เหมือน ๆ กัน หรืออาจจะเป็นในลักษณะที่กลับกัน (Negative relation) สำหรับความสัมพันธ์ของตัวแปรที่นำมาศึกษานั้นจะสูงหรือต่ำ ในบางครั้งจะขึ้นอยู่กับลักษณะของข้อมูลในแต่ละตัวแปรที่นำมาศึกษาด้วย โดยที่จะพิจารณาจากตัวเลขเท่านั้น แต่สำหรับเครื่องหมายบวก หรือลบของค่าสัมประสิทธิ์สหสัมพันธ์นี้ จะแสดงให้ทราบถึงทิศทางของความสัมพันธ์นั้นว่าจะเป็นในลักษณะใด คือถ้าตัวแปรทั้งสองนี้อาจจะมีความสัมพันธ์ตามกัน (+) หรือสหสัมพันธ์ตรงกันข้าม (-) แต่ไม่สามารถนำมาบวก

ลบกันตามเครื่องหมายที่ปรากฎ ขนาดของค่าสหสัมพันธ์นี้ไม่สามารถตีความหมายเหมือนค่าร้อยละได้ ตัวเลขที่ใช้ในการคำนวณค่าสัมประสิทธิ์สหสัมพันธ์นี้ มักจะได้มาจากกลุ่มตัวอย่างที่ทำการศึกษาเท่านั้น การที่จะอ้างอิงไปถึงประชากรว่าตัวแปรทั้งคู่ที่ศึกษาจะมีความสัมพันธ์กันหรือไม่นั้น ขึ้นอยู่กับความมีนัยสำคัญของค่าสัมประสิทธิ์สหสัมพันธ์ที่คำนวณได้ โดยจะมีการกำหนดให้ดังนี้ $H_0 : \rho = 0$ และถ้าสมมติฐานศูนย์ (Null hypothesis) นั้น ถูกปฏิเสธย่อมแสดงว่า ค่าสัมประสิทธิ์สหสัมพันธ์ของประชากรนี้ไม่เท่ากับ 0

สำหรับการคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์นั้น จะต้องคำนึงถึงข้อตกลงเบื้องต้นของค่าสัมประสิทธิ์สหสัมพันธ์ที่จะใช้คำนวณ อีกทั้งยังจะต้องใช้ให้เหมาะสมกับระดับของข้อมูลที่มีอยู่ มิฉะนั้นจะทำให้ผลสรุปที่ได้นั้นอาจไม่น่าเชื่อถือเท่าที่ควร ทั้งในด้านความแม่นยำ และความตรงของการวิจัยด้วย (Lindeman and et al., 1980)

สัมประสิทธิ์สหสัมพันธ์แบบฟี (Phi correlation coefficient or Fourfold point correlation) สามารถเขียนสัญลักษณ์แทนด้วย ϕ หรือ r_ϕ โดยที่ค่าสัมประสิทธิ์สหสัมพันธ์นี้จะเป็นการวัดความสัมพันธ์ระหว่างตัวแปรเป็นตัวแปรทวิวิภาคทั้งคู่ (Genuinely dichotomous variable) โดยที่ข้อตกลงเบื้องต้นของตัวแปรทวิวิภาคนี้จะไม่ต่อเนื่อง (Discrete) (Ferguson, 1959) หรือตัวแปรตัวหนึ่งอาจจะเป็นตัวแปรทวิภาค (Dichotomous) และอีกตัวหนึ่งอาจจะเป็นตัวแปรทวิวิภาคภายใต้การแจกแจงปกติ (Artificial dichotomous with underlying normal distribution) ซึ่งในทางปฏิบัตินั้นค่าสัมประสิทธิ์สหสัมพันธ์นี้จะใช้ได้กว้างขวางเมื่อตัวแปรทั้งสองถูกเห็นได้ชัดว่าไม่ต่อเนื่อง (Ferguson, 1959) ซึ่ง ลินเดอร์แมนและคณะ (Lindeman and et.al., 1980) ได้กล่าวว่าสัมประสิทธิ์สหสัมพันธ์แบบนี้อาจจะถูกใช้แสดงดีกรีของความเกี่ยวพัน (Degree of association) ระหว่างตัวแปรทวิวิภาคทั้งสอง หรือถูกใช้เป็นดัชนีสนับสนุนซึ่งแตกต่างไปจากสัมประสิทธิ์สหสัมพันธ์แบบเตตระคลอริค (Tetrachoric correlation coefficiant) ที่ว่าตัวแปรทั้งคู่จะต้องเป็นตัวแปรทวิวิภาคภายใต้การแจกแจงปกติ อย่างไรก็ตาม การ์บิน,เท็นจ์ และเวอร์เลจ (Garbin, Teng and Verlag, 1988) ได้กล่าวไว้ว่า สัมประสิทธิ์สหสัมพันธ์แบบฟีนี้ถือเป็นสูตรสหสัมพันธ์เพียร์สันในกรณีพิเศษ (Pearson correaltion formulas in special case) ซึ่งสอดคล้องกับเมเรนจ์ (Mehrens, 1967) ที่ว่าค่าสัมประสิทธิ์สหสัมพันธ์แบบนี้ คือ อนุพันธ์ (Derivative) ของค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สัน โพรดักซ์ โมเมนต์ (Pearson's Product Moment Coefficient) นั่นเอง และในงานวิจัยของกิลฟอร์ดและเพอร์รี่ (Guilford and Perry, 1951) ได้กล่าวว่า ค่าสัมประสิทธิ์สหสัมพันธ์แบบฟีนี้ จะเป็นการประมาณค่าของค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน โดยที่สเกลของแต่ละตัวแปรทวิวิภาคนั้นจะมีค่าเป็น 0 หรือ 1 จะเห็นได้ว่าสัมประสิทธิ์สหสัมพันธ์แบบฟีนี้ อาจจะถูกพิจารณาว่าเป็นทั้งการวัดสหสัมพันธ์

(Correlation) หรือ ความเกี่ยวพัน (Association) นั้น จะขึ้นอยู่กับการแบ่งชั้นของตัวแปร (Categories of the variables) ที่ถูกจัดอันดับหรือไม่ถูกจัดอันดับ (Lindeman and et al., 1988) ส่วนใหญ่ข้อมูลที่ใช้ในการคำนวณค่าสัมประสิทธิ์สหสัมพันธ์แบบนี้จะถูกจัดให้อยู่ในลักษณะของความถี่ ในตารางการณ์จรขนาด 2 x 2 มิติ ซึ่งสามารถเขียนแสดงเป็นภาพประกอบการอธิบายได้ดังนี้

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

สำหรับสูตรที่ใช้ในการคำนวณสัมประสิทธิ์สหสัมพันธ์แบบฟีที่นิยมใช้กันในปัจจุบันนี้จะคำนวณโดยตรงจากความถี่ในแต่ละเซลของในตารางการณ์จรและผลรวมข้างของตาราง (marginal) ดังสมการข้างล่างนี้

$$r_\phi = \frac{bc - ad}{\sqrt{(a + b)(a + c)(b + d)(c + d)}} \qquad (31)$$

(Glass and Hopkin, 1984)

โดยที่ a, b, c และ d แทนความถี่ในแต่ละเซลของตารางการณ์จร (fourfold table) เมเรนซ์ (mehrens, 1967) โกซ์เอาท์ (Kohout, 1974) และ ลินเดอร์แมนและคณะ (Lindeman and et.al., 1980) ได้กล่าวว่าจากสมการที่ 31 นี้ ถือว่าเป็นอนุพันธ์ของสัมประสิทธิ์สหสัมพันธ์เพียร์สัน โพรดักซ์ โมเมนต์ โดยที่ค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สันที่ถูกคำนวณบนมาตราบัญญัติ โดยที่ข้อมูลเป็น Dichotomous นี้จะถูกเรียกว่า สัมประสิทธิ์สหสัมพันธ์แบบฟี (Glass and Stanley, 1970) และจากสูตรที่ว่า

$$r_{xy} = \frac{n\sum XY - \sum X \sum Y}{\sqrt{[n\sum X^2 - (\sum X)^2][n\sum Y^2 - (\sum Y)^2]}} \qquad (32)$$

โดยที่ตัวแปร X และ ตัวแปร Y เป็นตัวแปรทวิวิภาค (Dichotomous variables) และ ΣX, ΣY, ΣX^2, ΣY^2 และ ΣXY นั้น สามารถแสดงในเทอมของความถี่แต่ละเซลในตารางการณ์จรได้ดังนี้คือ

$$\Sigma XY = b, \quad \Sigma X \text{ และ } \Sigma X^2 = b+d, \quad \Sigma Y \text{ และ } \Sigma Y^2 = a+b$$

ซึ่งถ้านำค่าที่ได้เหล่านี้ไปแทนในสมการที่ 32 จะสามารถเขียนได้ในเทอมของ a's, b's, c's และ d's ได้ดังสมการข้างล่างนี้

$$r_\phi = \frac{nb - (b+d)(a+b)}{\sqrt{[n(b+d) - (b+d)^2][n(a+b) - (a+b)^2]}} \qquad (33)$$

แต่ n มีค่าเท่ากับ a+b+c+d ดังนั้นนำ a+b+c+d ไปแทน n ในสมการที่ 33

$$r_\phi = \frac{(a+b+c+d)b - ab - b^2 - ad - bd}{\sqrt{(b+d)(a+c)(a+b)(c+d)}} \qquad (34)$$

จากสมการที่ 34 นี้ จะสามารถเขียนเป็นสูตรสัมประสิทธิ์สหสัมพันธ์แบบฟีให้ง่ายและสะดวกในการนำไปใช้ดังสมการข้างล่างนี้

$$r_\phi = \frac{bc - ad}{\sqrt{(a+b)(a+c)(b+d)(c+d)}} \qquad (35)$$

ในปี คศ. 1901 คาร์ล เพียร์สัน ได้คิดค้นสูตรที่ใช้ในการคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์แบบฟีที่คำนวณได้จากความถี่ ดังสมการที่ 31 ในรายงานที่เผยแพร่ใน Philosophical Transactions of Royal Society of London ซึ่งสหสัมพันธ์ระหว่างตัวแปรจำนวนมากไม่สามารถถูกวัดเป็นปริมาณ (Measured quantitatively) และค่าสัมประสิทธิ์สหสัมพันธ์ที่คำนวณได้จากสมการที่ 31 นี้ จะมีผลลัพธ์เหมือนกับสมการที่ 35 ทั้งนี้เนื่องจากสมการที่ 31 และสมการที่ 35 นี้ เป็นพีชคณิตที่มีค่าเท่ากัน (Algebraically equivalent) (Glass and Stanley, 1970)

$$r_{\phi} = \frac{p_{xy} - p_x p_y}{\sqrt{p_x q_x p_y q_y}} \tag{36}$$

โดยที่ p_x แทน สัดส่วนของบุคคลที่ได้คะแนน 1 บนตัวแปร x

q_x แทน สัดส่วนของบุคคลที่ได้คะแนน 0 บนตัวแปร x ซึ่งจะเท่ากับ $1 - p_x$

p_y แทน สัดส่วนของบุคคลที่ได้คะแนน 1 บนตัวแปร y

q_y แทน สัดส่วนของบุคคลที่ได้คะแนน 0 บนตัวแปร y ซึ่งจะเท่ากับ $1 - p_y$

p_{xy} แทน สัดส่วนของบุคคลที่ได้คะแนน 1 บนตัวแปร x และตัวแปร y

สำหรับค่าสัมประสิทธิ์สหสัมพันธ์แบบฟีที่คำนวณจากสมการที่ 31 นี้ จะมีเครื่องหมายเป็นบวก ก็ต่อเมื่อความถี่บนไดออกอนอล bc (bc diagonal) มีค่าใหญ่กว่าไดออกอนอล ad (ad diagonal) ซึ่งในทางกลับกันนั้นถ้าค่า ad นั้นมีค่าใหญ่กว่า bc จะทำให้ค่าสัมประสิทธิ์สหสัมพันธ์นี้มีเครื่องหมายเป็นลบ ซึ่งเครื่องหมายของค่าสัมประสิทธิ์สหสัมพันธ์แบบฟีนี้จะมีความสัมพันธ์แปรผันโดยตรงกับค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สัน (r_{xy}) และค่าสัมประสิทธิ์สหสัมพันธ์ดังกล่าวนี้ ไม่จำเป็นต้องใช้กับตารางการณ์จรขนาด 2 x 2 มิติ และภายใต้เงื่อนไขบางอย่างค่าสัมประสิทธิ์สหสัมพันธ์แบบนี้สามารถมีค่ามากกว่า 1.00 ได้ถูกคำนวณจากตารางการณ์จรขนาดใหญ่ (Large contigence table) แต่สัมประสิทธิ์สหสัมพันธ์นี้ไม่เหมาะสมสำหรับตารางการณ์จรขนาดใหญ่ สำหรับกรณีที่แถวและคอลัมน์ของตารางการณ์จรขนาดใหญ่นี้เป็นตัวแทนของการจัดลำดับชั้น (Order categories) สัมประสิทธิ์สหสัมพันธ์ที่เหมาะสมในกรณีนี้คือ เคนดอล ทาว์น (Kendal's Tau : τ) หรือ โซเมอร์ดี (Somer's d) (Kohout, 1974)

สำหรับสูตรพื้นฐานที่ใช้สำหรับการคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์แบบฟี นั่นคือ

$$r_{\phi} = \frac{\alpha\nu - \beta\gamma}{\sqrt{pqp'q'}} \tag{37}$$

p, q, p' และ q' เป็นสัดส่วนของมาร์จินอล (Marginal) ในตารางการณ์จรโดยที่ α, ν, β, และ γ เป็นความถี่ในแต่ละเซล โดยที่จะต้องมีความถี่ตรงกับรูปแบบของสัดส่วนที่แสดงในตารางการณ์จรข้างล่างนี้

ตัวแปรที่ 1

		0	1	Total
ตัวแปรที่ 2	1	β	α	p
	0	ν	γ	q
	Total	q'	p'	1.00

(Guilford and Fruchther, 1978)

จากสมการที่ 37 นี้ ถ้าค่า $p' = q' = 0.50$ สูตรที่ใช้ในการคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์นี้จะคำนวณได้จากสัดส่วนดังจะปรากฏตามสูตรดังนี้

$$r_{\phi} = \frac{\alpha - \beta}{\sqrt{pq}} \qquad (38)$$

สัมประสิทธิ์สหสัมพันธ์แบบฟีนี้จะมีความสัมพันธ์ใกล้ชิดกับไคสแควร์ (Chi Square) ซึ่งสามารถที่จะนำมาประยุกต์ใช้กับสถานการณ์ต่าง ๆ ได้อย่างกว้างขวาง สำหรับตารางการณ์จรที่มีขนาด 2 x 2 มิตินี้ ค่าของไคสแควร์จะมีค่าเท่ากับฟังก์ชั่นของค่าสัมประสิทธิ์สหสัมพันธ์แบบฟี นั่นคือ

$$\chi^2 = Nr_{\phi}^2 \qquad : df = 1 \qquad (39)$$

และค่าสัมประสิทธิ์สหสัมพันธ์แบบฟี จะเป็นอนุพันธ์ (Derived) จากไคสแควร์ดังสมการข้างล่างนี้

$$r_{\phi} = \sqrt{\frac{\chi^2}{N}} \qquad (40)$$

(Guilford and Fruchter,1978)

r_{ϕ} = Phi coefficient

χ^2 = ค่าที่คำนวณได้จากไคสแควร์

N = จำนวนของกลุ่มตัวอย่าง (Total number of Cases)

จากสูตรในสมการที่ 39 นี้ จะเป็นสูตรในการทดสอบนัยสำคัญของสัมประสิทธิ์สหสัมพันธ์แบบฟี ซึ่งในการทดสอบสมมติฐานศูนย์ (Null hypothesis) $H_o: \rho = 0$ ซึ่งภายใต้สมมติฐานศูนย์ (H_o) การแจกแจงของฟี (ϕ) จะมีลักษณะการแจกแจงเป็นไคสแควร์นี้ มีชั้นแห่งความเป็นอิสระ (df) เท่ากับ 1 โดยที่ค่าสัมประสิทธิ์สหสัมพันธ์แบบฟีนี้จะหาค่าได้จากไคสแควร์ที่ประยุกต์กับตารางการณ์จรขนาด 2 x 2 มิติเท่านั้น การปฏิเสธสมมติฐานศูนย์นั้นจะทำได้ก็ต่อเมื่อค่าไคสแควร์ที่คำนวณได้จากสูตรจะมีค่ามากกว่าค่าไคสแควร์ที่เปิดตาราง ซึ่งกิลฟอร์ดและฟรัตเตอร์ (Guilford and Fruchter, 1978) ได้กล่าวไว้ว่า สำหรับค่าไคสแควร์ที่คำนวณได้จากตารางการณ์จรนั้นมีนัยสำคัญทางสถิติแล้ว จะทำให้ค่าสัมประสิทธิ์สหสัมพันธ์แบบฟีนั้นจะมีนัยสำคัญทางสถิติด้วย หรืออาจจะทดสอบนัยสำคัญของสัมประสิทธิ์สหสัมพันธ์แบบนี้ โดย

$$Z = \frac{\phi}{1 / \sqrt{N}} \qquad (41)$$

เมื่อ N เป็นขนาดของกลุ่มตัวอย่างที่มีขนาดใหญ่ (N > 100) (Thorndike, 1978) ซึ่งค่าไคสแควร์นั้น จะถูกใช้โดยสูตรที่คล้ายคลึงกันสำหรับสูตรในสมการที่ 40 แต่จะไม่นำเข้าไปพัวพันกับค่าสัมประสิทธิ์สหสัมพันธ์แบบฟี

$$\chi^2 = \sum_{I=1}^{2} \sum_{j=1}^{2} \left(\frac{o_{ij} - e_{ij}^2}{e_{ij}} \right) \qquad (42)$$

โดยที่ o_{ij} แทน ความถี่ที่สังเกตได้ในเซล i (The observed frequency in cell i)

e_{ij} แทน ความถี่ที่คาดหวังในเซล i (The expected frequency in cell i)

เมื่อตัวแปร X และตัวแปร Y เป็นการแบ่งชั้นโดยการจัดอันดับ ช่วงของค่าสัมประสิทธิ์สหสัมพันธ์ (r_ϕ) จะมีค่าตั้งแต่ -1 ถึง 1 อย่างไรก็ตามถ้าตัวแปร X และตัวแปร Y เป็นการแบ่งชั้นโดยไม่จัดอันดับ ยกตัวอย่างเช่นความสัมพันธ์ระหว่างสีผมและสีตานั้น ถือว่าเป็นการวัดดีกรีของความสัมพันธ์เท่านั้น (Measure of degree of assoiciation) ซึ่งจะมีช่วงของ

ค่าสัมประสิทธิ์นี้ตั้งแต่ 0 ถึง 1 นั่นคือ ไม่มีความสัมพันธ์กันเลยถึงมีความสัมพันธ์กันในทางบวก

ขีดจำกัด(Limitations) ของขนาดของสัมประสิทธิ์สหสัมพันธ์แบบฟี

ถึงแม้ว่าสัมประสิทธิ์สหสัมพันธ์แบบฟีนั้น จะสามารถมีค่าตั้งแต่ -1 ถึง +1 ภายใต้เงื่อนไขที่แน่นอน ค่าสัมประสิทธิ์สหสัมพันธ์แบบนี้จะมีค่าเท่ากับ 1 ก็ต่อเมื่อ ค่าเฉลี่ยของ p และ p' มีค่าเท่ากัน และถ้าความแตกต่างระหว่างค่าเฉลี่ยทั้งสองมีมาก จะลด (reduction) ขนาดของสัมประสิทธิ์สหสัมพันธ์แบบฟี

ข้อกำหนดในการใช้สัมประสิทธิ์สหสัมพันธ์แบบฟี (Requirement for the use of the Phi coefficient correlation)

1. เป็นข้อมูลนามบัญญัติ (Nominal data) ซึ่งเป็นข้อมูลความถี่ (frequency data) เท่านั้น

2. ข้อมูลในตารางขนาด 2 x 2 มิติ จะสามารถจัดลงในรูปแบบของตารางขนาด 2 x 2 มิติ สัมประสิทธิ์สหสัมพันธ์แบบฟีนี้ไม่เหมาะสำหรับตารางการณ์จรที่มีขนาดใหญ่กว่า 2 x 2 มิติ

3. กลุ่มตัวอย่างเป็นแบบสุ่ม (Random sampling) ในการทดสอบนัยสำคัญของสัมประสิทธิ์สหสัมพันธ์แบบฟีนั้นจะต้องสุ่มอย่างมีหลักเกณฑ์ (Random basis) จากประชากรขนาดใหญ่ (Levin, 1983)

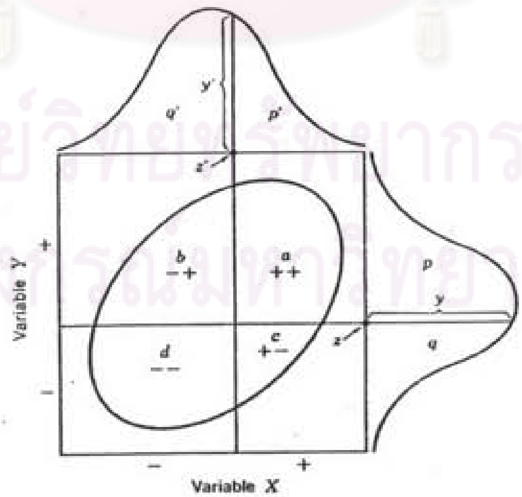
สัมประสิทธิ์สหสัมพันธ์แบบเตตระคลอริค (Tetrachoric Correlation Coefficient)

สัมประสิทธิ์สหสัมพันธ์เตตระคลอริคนี้ สามารถเขียนแทนด้วยสัญลักษณ์ $\mathbf{r_t}$ หรือ $\mathbf{r_{tet}}$ โดยที่สหสัมพันธ์ของตัวแปรทั้งสองนี้จะเป็นตัวแปรทวิวิภาคภายใต้การแจกแจงปกติ (Artificial dichotimied with underlying normal distribution) ซึ่งจะมีข้อตกลงเบื้องต้นเกี่ยวกับตัวแปรทั้งคู่นั้นจะต้องต่อเนื่อง (Continuous) มีการแจกแจงเป็นปกติ (Bivariate normal distribution) และตัวแปรเหล่านี้จะต้องมีความสัมพันธ์เชิงเส้นตรง (Linearly related variable) (Guilford and Fruchter, 1978) ซึ่งค่าสัมประสิทธิ์สหสัมพันธ์แบบเตตระคลอริคนี้ จะใช้วัดความสัมพันธ์ระหว่างคะแนนบนแบบสอบรายข้อ โดยที่สูตรที่เป็นพื้นฐานในการคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์แบบเตตระคลอริคนี้ได้พัฒนาโดย คาร์ล เพียร์สัน (Karl Pearson) ในปี คศ. 1901 ซึ่งจะนำไปสู่อนุกรมที่ไม่จำกัด (Infinite series) ซึ่งจะมีสูตรดังนี้

$$r_t + r_t^2 + \frac{zz'}{2} + r_t^3\frac{(z^2 - 1)(z'^2 - 1)}{6} + \ldots = \frac{ad - bc}{yy'N^2} \qquad (41)$$

(Guilford and Fruchter, 1978)

ซึ่งสัญลักษณ์ต่าง ๆ ในสมการที่ 41 นี้จะสามารถดูได้จากแผนภูมิข้างล่างนี้



แผนภาพที่ 3 การแจกแจงแบบปกติในกรณีที่แบ่งตัวแปร X และตัวแปร Y ในตารางการณ์จร และแสดงให้เห็นตัวคงที่ที่ปรากฏในสมการสำหรับสูตรคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์แบบเตตระคลอริค

โดยที่ a,b,c,d แทน ความถี่แต่ละเซลในตารางการณ์จร (fourfold table)

a และ d แทน ความถี่ที่มีเครื่องหมายเหมือนกัน (Yes-Yes) และ (No-No) ตามลำดับ

b และ c แทน ความถี่ที่มีเครื่องหมายต่างกัน (No-Yes) และ (Yes-No) ตามลำดับ

p'และ q' แทน สัดส่วนของการตอบคำถามที่ 1 ซึ่งเป็นการแบ่งตามลำดับชั้น (Two - category distribution of all responses to question 1)

p และ q แทน สัดส่วนการตอบคำถามที่ 2 ซึ่งจะเป็นการแจกแจงของผู้ตอบทั้งหมด

z และ z' แทน การวัดคะแนนมาตรฐาน (Standard score measurements) บนเส้นฐาน (Base line) ของโค้งการแจกแจงปกติ(Unit normal distribution curve) ที่จุดของการแบ่งกรณีในการแจกแจงทั้งสอง

y และ y' แทน ออร์ดิเนต(Ordinates)เช่นเดียวกับ z และ z'ในการแจกแจงปกติ (Unit Normal Distribution)

ซึ่งจะเห็นได้ว่าสูตรที่แท้จริงที่ใช้ในการคำนวณหาสัมประสิทธิ์สหสัมพันธ์นี้ยาว และซับซ้อน โดยที่สูตรนี้จะอยู่ในเทอมของอนุกรมไม่จำกัด ซึ่งก่อให้เกิดความยุ่งยากมากในการคำนวณ กลาสและฮอบกิ้นส์ (Glass and Hopkins, 1984) ได้เสนอแนะให้ใช้สูตรในการประมาณค่าสัมประสิทธิ์สหสัมพันธ์นี้คือ

$$r_{tet} = \frac{ad - bc}{U_x U_y n_o^2} \qquad (42)$$

(Glass and Hopkins, 1984)

โดยที่ a,b,c,d แทน ความถี่แต่ละเซลในตารางการณ์จร

U_xและU_y แทน ออร์ดิเนตของการแจกแจงปกติ(Unit normal distribution)ที่ P_xและ P_y

P_xและP_y แทน สัดส่วนของผู้ที่ตอบ “ 1 ” บนการวัดตัวแปร X และตัวแปร Y ตามลำดับ

n_o แทน จำนวนกลุ่มตัวอย่างที่ใช้ในงานวิจัย (The total number of cases)

สำหรับการใช้สูตรที่ 42 นี้จะต้องมีขนาดของกลุ่มตัวอย่างขนาดใหญ่นั่นคือ ควรมีจำนวน 200 หรือมากกว่านั้น สำหรับวิธีการประมาณค่าสัมประสิทธิ์สหสัมพันธ์แบบเตตระคลอริค (Method of estimating the tetrachoric r) นี้ สามารถใช้สูตรโคซายน์พาย (Cosine - Pi formula) ในการประมาณค่าสัมประสิทธิ์สหสัมพันธ์แบบเตตระคลอริค ในรูปแบบทางคณิตศาสตร์ (Mathematical form) ในเทอมของเรเดียน (Terms of radians) ดังนี้คือ

$$r_{tet} = \cos\left(\pi \frac{\sqrt{bc}}{\sqrt{ad} + \sqrt{bc}}\right) \qquad (43)$$

จากสมการที่ 43 นี้ ผู้วิจัยสามารถแทนค่า π ด้วย 180° ซึ่งจะอยู่ในเทอมของดีกรี (Term of degrees) ซึ่งจะก่อให้เกิดความสะดวกในการคำนวณมากขึ้น

$$r_{tet} = \cos\left(\frac{180^\circ \sqrt{bc}}{\sqrt{ad} + \sqrt{bc}}\right) \qquad (44)$$

จากสมการที่ 44 นี้ เมื่อนำ bc ไปหารทั้งเศษและส่วน จะทำให้สูตรที่ใช้ในการคำนวณสะดวกยิ่งขึ้น ซึ่งนับว่าเป็นสูตรโคซายน์ พาย ที่อยู่ในรูปง่าย ที่คำนวณได้จากความถี่ในแต่ละเซลของตารางการณ์จร

$$r_{tet} = \cos\left(\frac{180^\circ}{1 + \sqrt{ad/bc}}\right) \qquad (45)$$

ในกรณีที่สหสัมพันธ์มีค่าเป็น +1 (Perfect positive correlation) เมื่อ bc = 0 โดยที่ $r_t = \cos 0^\circ$ ซึ่งค่าโคซายน์ของมุม 0 จะมีค่าเป็น 1 ในทางกลับกันในกรณีที่สหสัมพันธ์มีค่าเป็น −1 (Perfect negative correlation) เมื่อ ad = 0 ซึ่ง ad/bc = 0 และ $r_t = \cos 180^\circ$ ซึ่งค่าโคซายน์ของมุม 180 จะมีค่าเป็น −1 และถ้าในกรณีที่ ad = bc จะทำให้ ad/bc = 1

และค่า $r_t = \cos 90^\circ$ ซึ่งค่าโคซายน์ของมุม 90 จะเป็น 0 และถ้ามุมนี้มีขนาดใหญ่กว่า 90° นั้น จะทำให้สหสัมพันธ์มีค่าเป็นลบ (Ferguson, 1959)

สำหรับสูตรที่ใช้ในการประมาณค่าสัมประสิทธิ์สหสัมพันธ์แบบเตตระคลอริค เมื่อตัวแปรทั้งสองเป็นการแบ่งโดยการประมาณค่าที่ค่ามัธยฐาน(Spilt approximately at the median) ดังสมการข้างล่างนี้

$$r_{tet} = \left[\sin \; 90^\circ \left(\frac{a+d-b-c}{n} \right) \right] \qquad (46)$$

แต่ถ้าการแบ่งไม่เข้าใกล้ค่ามัธยฐานบนตัวแปรตัวใดตัวหนึ่งหรือ ทั้งสองตัวแปรนี้ไม่ถูกแบ่งให้เข้าใกล้ค่ามัธยฐาน ก็สมควรที่จะใช้สูตรในสมการที่ 46 มากกว่า (Lindeman and et.al., 1980) ซึ่งสอดคล้องกับ เฟอร์กูสัน (Ferguson, 1959) ที่ว่าสูตรโคซายน์-พายน์ จะใช้ประมาณค่าสัมประสิทธิ์สหสัมพันธ์แบบเตตระคลอริคได้ดีที่สุด ก็ต่อเมื่อการแบ่งของตัวแปรทั้งสอง (Division of the two variables) เท่ากัน นั่นคือ $p = q = 0.5$

ค่าสัมประสิทธิ์สหสัมพันธ์แบบเตตระคลอริคนี้ จะมีความเที่ยงตรงน้อยกว่า (less reliable) ค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน โพรดักซ์ โมเมนต์ อย่างน้อย 50 % ของตัวแปรที่มากกว่า (Guliford and Fruchter, 1978) ซึ่งค่าสัมประสิทธิ์สหสัมพันธ์นี้จะมีคาความเที่ยงมากที่สุด เมื่อ

1. N มีขนาดใหญ่ (N is large as is true of all statistics)
2. เมื่อค่าสัมประสิทธิ์สหสัมพันธ์แบบเตตระคลอริคนั้นมีขนาดใหญ่ (r_t is large as is true of other r'.)
3. การแบ่งในสองกลุ่ม (Two categories) นั้น จะเข้าใกล้ค่ามัธยฐาน (Medians)

ข้อจำกัดในการใช้สูตรโคซายน์-พาย(Limitations to the Cosine - Pi Formula)

สำหรับสูตรที่ 45 นี้ จะประมาณค่าสัมประสิทธิ์สหสัมพันธ์แบบเตตระคลอริคได้ใกล้เคียงที่สุด เมื่อตัวแปรทั้งสองเป็น Dichotomous ที่ถูกแบ่งที่มัธยฐาน (Guilford and Fruchter, 1978) เมื่อ $\rho = 0$ ในประชากรปกติสองตัวแปร ซึ่งเป็นข้อตกลงเบื้องต้นของสัมประสิทธิ์สหสัมพันธ์แบบเตตระคลอริค สำหรับการแจกแจงของกลุ่มตัวอย่างของสัมประสิทธิ์สหสัมพันธ์นี้ เป็นการประมาณค่าปกติสำหรับกลุ่มตัวอย่างที่มี $n = 30$ หรือมากกว่านั้น

(Moderately large n) และเมื่อจุดแบ่งของตัวแปรทั้งสอง (Division of the two variables) ยิ่งแตกต่างกันมาก จะทำให้ค่าสัมประสิทธิ์สหสัมพันธ์แบบเตตระคลอริคที่ได้นี้ มีค่าสูงกว่าความเป็นจริง (Over estimate) นอกจากนี้การคำนวณที่ใช้สูตรโคซายน์ - พาย นี้ จะมีความคลาดเคลื่อนในการสุ่มมาก จึงต้องใช้กลุ่มตัวอย่างขนาดใหญ่ ปกติจะใช้ขนาดของกลุ่มตัวอย่าง 2 เท่าของที่ใช้ในการคำนวณด้วยสูตรสัมประสิทธิ์สหสัมพันธ์เพียร์สัน โพรดักซ์ โมเมนต์

สูตรที่ใช้ในการคำนวณค่าความคลาดเคลื่อนมาตรฐานที่ใช้ในการทดสอบสมมติฐานของค่าสัมประสิทธิ์สหสัมพันธ์นี้ (Standard error for testing the hypothesis of zero Correlation) จะถูกประมาณค่าโดย

$$S_{rt} = \frac{pp'qq'}{yy'\sqrt{N}} \tag{47}$$

(Guilford and Fruchter, 1978)

โดยที่ p' s และ q' s เป็นสัดส่วนที่มีพื้นฐานบนความถี่ในแต่ละเซลดังแสดงในตารางข้างต้น สำหรับการทดสอบนัยสำคัญของค่าสัมประสิทธิ์สหสัมพันธ์แบบเตตระคลอริคในประชากรนั้น จะตั้งสมมติฐานว่า Ho : $\rho = 0$ นั้น คือ

$$Z = \frac{r_{tet}}{S_{r_{tet}}} \tag{48}$$

(Lindeman, Merenda and Gold, 1980)

ดีกรีของความเที่ยงนั้นในค่าสัมประสิทธิ์สหสัมพันธ์แบบเตตระคลอริคนั้น จะเหมือนกับดีกรีของค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สัน โพรดักซ์ โมเมนต์ สำหรับการประมาณค่าดีกรีของสหสัมพันธ์โดยค่าเฉลี่ยของค่าสัมประสิทธิ์สหสัมพันธ์แบบเตตระคลอริคนั้น กิลฟอร์ด และฟรัคเตอร์ (Guilford and Fruchter, 1978) ได้แนะนำว่าขนาดของกลุ่มตัวอย่าง (N) อย่างน้อยควรจะเป็น 200 และ ถ้าจะให้ดีควรจะเป็น 300 หรือมากกว่านั้น ซึ่งสอดคล้องกับลินเดอร์แมนและคณะ (Lindeman and et.al., 1980) ที่กล่าวไว้ว่า ผู้วิจัยควรระมัดระวังในการแปลความหมายของค่าสัมประสิทธิ์สหสัมพันธ์แบบเตตระคลอริคนี้ ถ้ากลุ่มตัวอย่างมีขนาดเล็ก และไม่แนะนำให้ใช้

ค่าสัมประสิทธิ์สหสัมพันธ์นี้ ถ้ามีขนาดของกลุ่มตัวอย่างใหญ่พอ (อย่างน้อย 2 เท่า หรือมากกว่านั้น) และกิลฟอร์ด และ ฟรัคเตอร์ (Guilford and Fruchter, 1978) ได้กล่าวเพิ่มเติมว่า สำหรับในกลุ่มตัวอย่างที่มีขนาดเล็กกว่านี้ โดยที่จำนวนของกลุ่มตัวอย่างมี (N) น้อยกว่า 100 ค่าสัมประสิทธิ์สหสัมพันธ์แบบเตตระคลอริคนี้จะสามารถใช้ในการทดสอบสมมติฐานศูนย์ได้ (Test null hypothesis)

ข้อหลีกเลี่ยงในการใช้ค่าสัมประสิทธิ์สหสัมพันธ์แบบเตตระคลอริคนั้นก็คือ ขนาดของกลุ่มตัวอย่างในตารางแจกแจงความถี่นั้น ไม่ควรมีขนาดของความถี่ในตารางการณ์จรแตกต่างกันมาก เช่น 95 - 5 หรือ 90 - 10 ทั้งนี้เนื่องจากความถี่ในแต่ละเซลที่แตกต่างกันมากของตารางการณ์จรจะทำให้ค่าความคลาดเคลื่อนมาตรฐานเพิ่มขึ้นในสถานการณ์นี้ โดยเฉพาะอย่างยิ่งควรพยายามหลีกเลี่ยงการประมาณค่าสัมประสิทธิ์สหสัมพันธ์แบบนี้ เมื่อมีความถี่เป็น 0 ในตารางการณ์จรเพียงเซลเดียว ซึ่งจะมีผลทำให้ค่าสัมประสิทธิ์สหสัมพันธ์ที่คำนวณได้ไม่ตรงกับความเป็นจริง นอกจากนี้แล้ว การ์บิน,เทนจ์ และเวอร์เลจ (Garbin, Teng and Verlag, 1988) ได้กล่าวว่า ค่าสัมประสิทธิ์สหสัมพันธ์แบบเตตระคลอริคจะใหญ่กว่าค่าสัมประสิทธิ์สหสัมพันธ์แบบฟี ถ้าใช้ข้อมูลชุดเดียวกัน และค่าสัมประสิทธิ์สหสัมพันธ์แบบเตตระคลอริคจะมีช่วงตั้งแต่ -1 จนถึง +1 กลาสและสแตนเลย์ (Glass and Stanley, 1970) ได้กล่าวไว้ว่าค่าสัมประสิทธิ์สหสัมพันธ์แบบเตตระคลอริคจะดีกว่าสัมประสิทธิ์สหสัมพันธ์แบบฟีในด้านการวัดความสัมพันธ์ (Measure of relationship) เมื่อตัวแปรทวิวิภาคภายใต้การแจกแจงปกติ (Normal distributions underlying the dichotomies)

ตารางที่ 2 ความแตกต่างระหว่างการใช้ค่าสัมประสิทธิ์สหสัมพันธ์แบบฟีกับแบบเตตระคลอริค

	ค่าสัมประสิทธิ์สหสัมพันธ์ แบบฟี	ค่าสัมประสิทธิ์สหสัมพันธ์ แบบเตตระคลอริค
สัญลักษณ์ที่ใช้	r_{ϕ} หรือ ϕ	r_{tet} หรือ r_t
ตัวแปรที่ใช้ในการหาความสัมพันธ์	เป็นตัวแปรทวิวิภาคทั้งคู่ (True dichotomous) หรือตัวแปรตัวหนึ่งเป็นตัวแปรทวิวิภาคและอีกตัวหนึ่งเป็นตัวแปรทวิวิภาคภายใต้การแจกแจงแบบปกติ (Artificial dichotomies with underlying normal distribution)	เป็นตัวแปรทวิวิภาคภายใต้การแจกแจงแบบปกติ (Artificial dichotomies with underlying normal distribution)
สูตรที่ใช้ในการคำนวณ	$r_{\phi}=\dfrac{bc - ad}{\sqrt{(a+b)(a+c)(b+d)(c+d)}}$	$r_{tet} = \cos\left(\dfrac{180^{0}}{1+\sqrt{ad/bc}}\right)$
ปัญหาของการใช้ค่าสัมประสิทธิ์สหสัมพันธ์ในการวิเคราะห์ตัวประกอบ	1. มีจำนวนตัวประกอบมากกว่าปกติ (Difficulty factor) 2. ค่าความยากของข้อสอบนั้น มีผลต่อค่าสัมประสิทธิ์สหสัมพันธ์ที่คำนวณได้	1. เมื่อมีค่าการเดาเกิดขึ้นจะทำให้เกิด Spurious factor 2. ต้องใช้กลุ่มตัวอย่างขนาดใหญ่ เนื่องจากมีความคลาดเคลื่อนในการสุ่ม 3. เมื่อมีความถี่เป็น 0 ในตารางการณ์จรเพียงเซลเดียวจะมีผลทำให้ค่าสัมประสิทธิ์สหสัมพันธ์ที่คำนวณได้ไม่ตรงกับความเป็นจริง และถ้ามีค่า −1 หรือ+1 จะทำให้เมตริกซ์ที่ได้ไม่เป็น Positively definite

ตอนที่ 4 งานวิจัยที่เกี่ยวข้อง

พารี่ และ แมกซ์อาร์เดอร์ (Parry and McArdle, 1991) การศึกษาเปรียบเทียบการใช้ค่าสัมประสิทธิ์สหสัมพันธ์ที่คำนวณจากค่าสัมประสิทธิ์สหสัมพันธ์แบบฟีกับแบบเตตระคลอริค ด้วยวิธีการกำลังสองน้อยที่สุด (Least squares method) ทั้ง 2 วิธีในการวิเคราะห์ตัวประกอบซึ่งเป็นตัวแปรทวิวิภาค โดยแบ่งการสะกัดตัวประกอบออกเป็น 4 แบบดังนี้คือ

1. ULS-PHI ใช้เมตริกซ์สหสัมพันธ์ที่คำนวณจากค่าสัมประสิทธิ์สหสัมพันธ์แบบฟี และใช้วิธีการสะกัดตัวประกอบด้วย Unweighted least - squares method 2. ULS-TC ใช้เมตริกซ์สหสัมพันธ์ที่คำนวณจากค่าสัมประสิทธิ์สหสัมพันธ์แบบเตตระคลอริค และใช้วิธีการสะกัดตัวประกอบด้วย Unweighted least - squares method 3. LISCOMP ใช้เมตริกซ์สหสัมพันธ์ที่คำนวณจากค่าสัมประสิทธิ์สหสัมพันธ์แบบเตตระคลอริค และใช้วิธีการสะกัดตัวประกอบด้วย Weighted least - squares method 4. NOHARM ใช้เมตริกซ์สหสัมพันธ์แบบ Cross - product โดยใช้ค่าสัมประสิทธิ์สหสัมพันธ์แบบฟีบวกกับค่าเฉลี่ย (Phi coefficients plus means)และวิธีการสะกัดตัวประกอบใช้การประมาณค่าด้วย Unweighted least - squares method สำหรับตัวแปรทวิวิภาค สำหรับการศึกษาครั้งนี้มีขนาดของกลุ่มตัวอย่าง, ค่าจุดตัด (Threshold values) และดูจากค่าเบี่ยงเบนของน้ำหนักตัวประกอบของกลุ่มตัวอย่างไปยังประชากรที่แตกต่างกัน การสะกัดตัวประกอบที่ถูกจัดกระทำขึ้นนั้น จะดูได้จากดัชนีความสอดคล้อง ซึ่งจะเปรียบเทียบส่วนเบี่ยงเบนของการประมาณค่าน้ำหนักตัวประกอบที่ได้จากกลุ่มตัวอย่างไปยังประชากร ปรากฏว่า วิธีการที่ 3 และวิธีการที่ 4 นั้นไม่แตกต่างกว่าวิธีการที่ 1 และวิธีการที่ 2 และวิธีการทั้งสี่นี้ไม่สามารถใช้ได้ดีในทุกสถานการณ์ ถึงแม้ว่าข้อมูลจะมีความเบ้สูงและมีขนาดของกลุ่มตัวอย่างขนาดเล็ก