บทที่ 2

วรรณคดีที่เกี่ยวข้อง

งานวิจัยนี้เป็นการประยุกต์ใช้โมเดลล็อกลิเนียร์ (log-linear model) กับการวิเคราะห์สาเหตุเพื่อศึกษาปัจจัยที่ส่งผลต่อระยะเวลาที่ใช้ในการศึกษาของมหาบัณฑิตทางสังคมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ผู้วิจัยจึงนำเสนอวรรณคดีที่เกี่ยวข้องโดยนำเสนอแยกเป็น 3 ตอน คือ ตอนที่ 1 ความรู้ทั่วไปเกี่ยวกับโมเดลล็อกลิเนียร์ กล่าวถึง ความเป็นมาและลักษณะทั่วไป การสร้างตารางการณ์จร การสร้างโมเดลล็อกลิเนียร์ การประมาณค่าพารามิเตอร์ การเลือกโมเดลล็อกลิเนียร์ การตรวจสอบความเหมาะสมของโมเดลล็อกลิเนียร์ และการประมาณค่าพารามิเตอร์ลอการิทึมของอัตราส่วนแต้มต่อ ตอนที่ 2 การประยุกต์โมเดลล็อกลิเนียร์ในการวิเคราะห์สาเหตุ (causal analysis) และตอนที่ 3 งานวิจัยที่เกี่ยวข้อง

ตอนที่ 1 ความรู้ทั่วไปเกี่ยวกับโมเดลล็อกลิเนียร์

การวิเคราะห์ข้อมูลประเภทกลุ่มทำได้ด้วยการนำข้อมูลจำแนกลงในตารางการณ์จร (contingency table) แต่เนื่องจากในอดีตนั้น ทฤษฎีทางสถิติและเทคนิคการคำนวณสำหรับการวิเคราะห์ข้อมูลประเภทกลุ่มมีจำกัด การวิเคราะห์ข้อมูลจึงทำได้เฉพาะตาราง 2 มิติ และวิเคราะห์ข้อมูลโดยใช้การทดสอบความกลมกลืนไค-สแควร์ (χ^2 goodness of fit) ซึ่งพัฒนาโดย Karl Pearson (1900) ต่อมามีนักสถิติพัฒนาการวิเคราะห์ข้อมูลมากกว่า 2 มิติพร้อม ๆ กัน แต่ระเบียบวิธีไม่ชัดเจนเพียงพอที่นักวิจัยใหม่ ๆ จะนำไปใช้ Kennedy (1983) กล่าวว่า นักสถิติหลายท่าน เช่น Birch, Bishop, Fienberg, Goodman, Grizzle, Williams และ Haberman พยายามพัฒนาการวิเคราะห์จนได้เป็นระบบการวิเคราะห์มากมายที่ออกแบบเฉพาะกับการวิเคราะห์ข้อมูลพร้อม ๆ กันจากตารางการณ์จรพหุมิติ (multidimensional contingency table) นักสถิติดังกล่าวได้เสนอผลงานในช่วงเวลาต่างกันจนเป็นระบบการวิเคราะห์ที่เรียกทั่วๆ ไปว่าการวิเคราะห์ตารางการณ์จรแบบล็อกลิเนียร์ (log-linear contingency table analysis) หรือการวิเคราะห์ล็อกลิเนียร์ (log-linear analysis)

การวิเคราะห์ล็อกลิเนียร์เป็นผลจากการบูรณาการวิเคราะห์ข้อมูล 3 แบบ คือ การวิเคราะห์ความแปรปรวน (analysis of variance) การวิเคราะห์การถดถอย (regression analysis) และการทดสอบความกลมกลืนแบบไค-สแควร์หรือการทดสอบภาวะสารูปสนิท

(chi-square goodness of fit) เนื่องจากการวิเคราะห์ล็อกลิเนียร์วิเคราะห์ข้อมูลจากตัวแปรมากกว่า 2 ตัวแปรขึ้นไปในตารางหลาย ๆ มิติได้พร้อม ๆ กัน จึงทำให้ทราบอิทธิพลหลัก (main effect) และอิทธิพลปฏิสัมพันธ์ (interaction effect) ของตัวแปร และมีโมเดลนำเสนอความสัมพันธ์ระหว่างตัวแปรที่ศึกษาได้เช่นเดียวกับการวิเคราะห์ความแปรปรวนและการวิเคราะห์การถดถอย โมเดลล็อกลิเนียร์จะคล้ายกับโมเดลการวิเคราะห์ความแปรปรวนต่างกันตรงที่โมเดลการวิเคราะห์ความแปรปรวนเป็นโมเดลที่ทำนายค่าเฉลี่ยของตัวแปรตามด้วยอิทธิพลของตัวแปรต้นและโมเดลเป็นแบบบวก (additive model) ส่วนโมเดลในการวิเคราะห์ล็อกลิเนียร์เป็นโมเดลที่ทำนายโอกาสในการเกิดค่าความถี่ที่คาดหวัง (expected frequency) ของตารางการณ์จรด้วยชุดอิทธิพลของตัวแปรต้นและโมเดลเป็นแบบคูณ (multiplicative model) เมื่อคำนวณค่าลอการิทึม (logarithm) ทำให้โมเดลแบบคูณอยู่ในรูปโมเดลแบบบวกและแสดงความสัมพันธ์แบบเส้นตรง (linear relationship) การเปลี่ยนรูปโมเดลดังกล่าวจึงเป็นที่มาของชื่อโมเดลล็อกลิเนียร์ (log-linear model) (Kennedy, 1983; Stevens, 1996)

อย่างไรก็ตาม การวิเคราะห์ความแปรปรวนและการวิเคราะห์ล็อกลิเนียร์ยังมีความแตกต่างอื่น ๆ อีกซึ่ง Kennedy (1983) และ Stevens (1996) ได้อธิบายไว้สรุปได้ดังนี้ 1) การวิเคราะห์ความแปรปรวนมีโมเดลสำหรับวิเคราะห์เพียงโมเดลเดียว แต่การวิเคราะห์ล็อกลิเนียร์มีโมเดลเพื่อการพิจารณาหลายโมเดล 2) การวิเคราะห์ล็อกลิเนียร์สามารถวิเคราะห์ข้อมูลได้ 2 ลักษณะ ลักษณะแรก วิเคราะห์ความสัมพันธ์แบบสมมาตร (symmetrical relationship) เป็นการศึกษาความสัมพันธ์ระหว่างตัวแปรโดยที่ไม่ระบุว่าตัวแปรใดเป็นตัวแปรตามและตัวแปรใดเป็นตัวแปรต้น ลักษณะที่สอง วิเคราะห์ความสัมพันธ์แบบอสมมาตร (asymmetrical relationship) เป็นการศึกษาความสัมพันธ์เมื่อกำหนดให้ตัวแปรหนึ่งเป็นตัวแปรตามได้รับอิทธิพลจากชุดของตัวแปรอื่นที่กำหนดให้เป็นตัวแปรต้น แต่การวิเคราะห์ความแปรปรวนทำได้เพียงลักษณะที่สองเท่านั้น 3) ในการวิเคราะห์ความแปรปรวน ตัวแปรตามมีลักษณะการวัดเป็นมาตรอันตรภาค (interval scale) หรือมาตรอัตราส่วน (ratio scale) ในขณะที่ตัวแปรต้นจะมีลักษณะการวัดแบบมาตรนามบัญญัติ (nominal scale) หรือมาตรอันดับ (ordinal scale) แต่ในการวิเคราะห์ล็อกลิเนียร์ ตัวแปรทุกตัวมีลักษณะการวัดแบบมาตรนามบัญญัติหรือมาตรอันดับ

การวิเคราะห์ล็อกลิเนียร์ทั้งการศึกษาแบบสมมาตร และแบบอสมมาตรมีกระบวนการวิเคราะห์เหมือนกัน กระบวนการวิเคราะห์แบ่งได้เป็น 6 ขั้นตอน ขั้นตอนแรก สร้างตารางการณ์จรและคำนวณความถี่ที่คาดหวังในแต่ละเซลล์ของตาราง ขนาดของตารางอาจเป็น 2 มิติ 3 มิติ 4 มิติหรือมากกว่า ขึ้นกับจำนวนตัวแปร ขั้นตอนที่สอง สร้างโมเดลล็อกลิเนียร์ (log-linear modeling) ขั้นตอนที่สาม ประมาณค่าพารามิเตอร์ (parameter estimates) พารามิเตอร์ในที่นี้คือ อิทธิพลของตัวแปรในแต่ละเทอม ขั้นตอนที่สี่ คัดเลือกโมเดล (models selection) ด้วยการทดสอบความกลมกลืน (fitting log-linear model) ขั้นตอนที่ห้า ตรวจสอบความเหมาะสมของโมเดล และขั้นตอนที่หก ประมาณค่าพารามิเตอร์ พารามิเตอร์ในที่นี้คือ ลอการิทึมของอัตราส่วนแต้มต่อ (log odds ratio) สำหรับการแปลผล ดังรายละเอียดแต่ละหัวข้อต่อไปนี้

1. การสร้างตารางการณ์จร และการคำนวณความถี่ที่คาดหวัง

1.1 การสร้างตารางการณ์จร (Creating Contingency Table)

การวิเคราะห์ล็อกลิเนียร์เริ่มด้วยการสร้างตารางการณ์จร ข้อมูลในตาราง คือ ความถี่ (frequency) ซึ่งได้มาจากการแจงนับหน่วยตัวอย่างที่สุ่มจากประชากรตามข้อตกลงว่าหน่วยตัวอย่างแต่ละหน่วยต้องถูกจำแนกลงในแต่ละเซลล์ได้อย่างอิสระ (independent) และจำแนกลงเพียงเซลล์เดียวเท่านั้น ดังนั้น ตารางการณ์จรที่ได้จึงเป็นการแจกแจงความถี่ในแต่ละเซลล์แบบไม่เกิดร่วม (mutually exclusive) และแต่ละเซลล์ต้องมีความถี่มากพอที่จะวิเคราะห์ ขนาดของตาราง (dimension) จะมีขนาดเท่าใดหรือจำนวนเซลล์มีกี่เซลล์นั้น ขึ้นกับจำนวนตัวแปรและจำนวนค่าสังเกตของตัวแปรนั้น ๆ เช่น การศึกษาตัวแปร 2 ตัวแปร คือ A และ B โดยตัวแปร A มี i ค่า ส่วนตัวแปร B มี j ค่า สร้างตารางการณ์จร 2 มิติ ขนาด i x j ได้ขนาด i แถว (rows) และ j สดมภ์ (columns) ทำให้เกิดเซลล์ ij เซลล์ อิทธิพลหลักของ A พิจารณาจากค่าความถี่ที่เป็นผลรวมของตารางแต่ละแถว และอิทธิพลหลักของ B พิจารณาจากความถี่ที่เป็นผลรวมของสดมภ์ ส่วนปฏิสัมพันธ์ของ A และ B พิจารณาจากความถี่ในเซลล์ ij การศึกษาตัวแปร 3 ตัวแปร คือ A, B และ C โดยที่ A มี i ค่า B มี j ค่า และ C มี k ค่า จะสร้างตารางการณ์จร 3 มิติ ขนาด i X j X k เป็นต้น ตารางเหล่านี้เป็นตารางที่วิเคราะห์ข้อมูลได้ทั้งความสัมพันธ์ระหว่างตัวแปรแบบสมมาตรและอสมมาตร

สำหรับกรณีตัวแปร A มี 2 ค่า คือ 1 และ 2 และตัวแปร B มี 3 ค่า คือ 1 , 2 และ 3 จะสร้างตารางการณ์จร 2 มิติ ขนาด 2 X 3 ได้ดังตารางที่ 1

ตารางที่ 1 แสดงลักษณะของตารางการณ์จร 2 มิติ ขนาด 2 X 3

A \ B	1	2	3	รวม
1				
2				
รวม				

จากที่กล่าวแล้วว่าโมเดลล็อกลิเนียร์เป็นโมเดลที่ทำนายค่าความถี่ที่คาดหวังด้วยอิทธิพลของตัวแปรที่ศึกษา ความถี่ที่คาดหวัง (expected frequency) หมายถึง หน่วยตัวอย่างที่มีโอกาสถูกแจงนับตามคุณลักษณะของค่าที่วัดในแต่ละเซลล์ของตารางมีค่าเท่ากับผลคูณของหน่วยตัวอย่างกับความน่าจะเป็นที่หน่วยตัวอย่างแต่ละหน่วยตกอยู่ในเซลล์ที่วัดได้ค่า i ของ A และค่า j ของ B ฉะนั้น ผู้วิจัยต้องทราบค่าความน่าจะเป็นที่หน่วยตัวอย่างจะปรากฏได้ในแต่ละเซลล์ก่อน เช่น การศึกษาตัวแปร 2 ตัวแปร คือ A และ B ดังตารางที่ 1 เมื่อสุ่มตัวอย่าง n คน นำข้อมูลมาจำแนกเป็นความถี่ในแต่ละเซลล์ ความถี่ในที่นี้เป็นจำนวนนับของหน่วยตัวอย่างที่ได้จากการเก็บรวบรวมข้อมูล เรียกความถี่นี้ว่า ความถี่สังเกตได้ (observed frequency) และผลการจำแนกความถี่สังเกตได้ ดังตารางที่ 2 หน้า 14 ตอนบน

เมื่อกำหนดสัญญลักษณ์ที่ใช้ดังนี้

i = สัญญลักษณ์แทนแถวในตาราง หรือระดับค่าของตัวแปร A และ i = 1 , 2

j = สัญญลักษณ์แทนสดมภ์ในตาราง หรือระดับค่าของตัวแปร B และ j = 1 , 2 , 3

f_{ij} = ความถี่สังเกตได้ (observed frequency) ในแถวที่ i สดมภ์ที่ j หรือความถี่สังเกตได้เซลล์ที่ ij (cell frequency)

$n_{i.}$ = ผลรวมความถี่สังเกตได้ (marginal observed frequency) ตามแถวที่ i $= \sum_{j=1}^{3} f_{ij}$

$n_{.j}$ = ผลรวมความถี่สังเกตได้ (marginal observed frequency) ตามสดมภ์ที่ j $= \sum_{i=1}^{2} f_{ij}$

n = ผลรวมความถี่สังเกตได้ทั้งหมด $= \sum_{i=1}^{2}\sum_{j=1}^{3} f_{ij}$

จะเห็นได้ว่าตารางแจกแจงข้อมูลที่ได้มีลักษณะเช่นเดียวกับตารางการณ์จรหรือการทำตารางไขว้ (cross classification) ในการวิเคราะห์ข้อมูลทางสถิติโดยทั่วไป เมื่อนักวิจัยต้องการศึกษาความสัมพันธ์ระหว่างตัวแปรหรือความเป็นอิสระระหว่างตัวแปรนั่นเอง

ค่าความถี่สังเกตได้ f_{11} , f_{12} , f_{13} ในแถวที่ 1 เป็นความถี่ที่เกิดจากอิทธิพลเนื่องจากตัวแปร B เช่นเดียวกับความถี่สังเกตได้ f_{21} , f_{22} , f_{23} ในแถวที่ 2 ถ้าความถี่สังเกตได้ทั้ง 3 ค่าในแต่ละแถวมีค่าเท่ากัน แสดงว่าไม่มีอิทธิพลจากตัวแปร B ในทำนองเดียวกันค่าความถี่ในสดมภ์ที่ 1 คือ f_{11} , f_{21} เป็นความถี่ที่เกิดจากอิทธิพลเนื่องมาจากตัวแปร A เช่นเดียวกับความถี่สังเกตได้ f_{12} , f_{22} หรือ f_{13} , f_{23} ในสดมภ์ที่ 2 และ 3 ตามลำดับ ถ้าความถี่ทั้ง 2 ค่าในแต่ละสดมภ์มีค่าเท่ากัน แสดงว่า ไม่มีอิทธิพลจากตัวแปร A

ในวิชาสถิติ นอกจากจะมีการจัดทำตารางการณ์จร แสดงการแจกแจงความถี่แล้ว ยังมีการนำเสนอตารางการณ์จรในรูปค่าความน่าจะเป็นด้วย การหาค่าความน่าจะเป็นที่นิยมใช้ คือ การหาค่าความน่าจะเป็นมีเงื่อนไข (conditional probability) ซึ่งหมายถึงการพิจารณาความถี่ในแต่ละเซลล์เทียบกับผลรวมความถี่ในแต่ละแถวหรือแต่ละสดมภ์ ค่าความน่าจะเป็นมีเงื่อนไขในที่นี้คำนวณได้จากอัตราส่วนของความถี่สังเกตได้ในเซลล์ ij กับผลรวมความถี่ในแถวที่ i ($p_{j/i}$) หรือ อัตราส่วนของความถี่ที่สังเกตได้ในเซลล์ ij กับผลรวมความถี่ในสดมภ์ที่ j ($p_{i/j}$) โดยทั่วไปเมื่อนักวิจัยกำหนดให้ตัวแปร A เป็นตัวแปรอิสระและตัวแปร B เป็นตัวแปรตาม นักวิจัยนิยมหาค่าความน่าจะเป็นมีเงื่อนไข โดยให้ผลรวมความถี่ในแต่ละแถวเป็นฐานในการคำนวณ คือ $p_{j/i}$ ค่าความน่าจะเป็นมีเงื่อนไขแสดงไว้ในตารางที่ 2 ตอนล่าง สูตรในการคำนวณความน่าจะเป็นมีเงื่อนไขมีดังต่อไปนี้

$$p_{j/i} = f_{ij}/n_{i.} \quad \text{หรือ} \quad p_{i/j} = f_{ij}/n_{.j}, \quad \text{โดยที่} \sum_{j=1}^{3} p_{j/1} = \sum_{j=1}^{3} p_{j/2} = 1 \text{ และ } \sum_{i=1}^{2} p_{i/1} = \sum_{i=1}^{2} p_{i/2} = 1$$

ตารางที่ 2 แสดงการแจกแจงข้อมูลลงบนตารางการณ์จร 2 มิติ ขนาด 2 X 3

การแจกแจงความถี่สังเกตได้				
A \ B	1	2	3	รวม
1	f_{11}	f_{12}	f_{13}	$n_{1.}$
2	f_{21}	f_{22}	f_{23}	$n_{2.}$
รวม	$n_{.1}$	$n_{.2}$	$n_{.3}$	n
การแจกแจงความน่าจะเป็น				
A \ B	1	2	3	รวม
1	$f_{11}/n_{.1}$ หรือ $f_{11}/n_{1.}$	$f_{12}/n_{.2}$ หรือ $f_{12}/n_{1.}$	$f_{13}/n_{.3}$ หรือ $f_{13}/n_{1.}$	1.00
2	$f_{21}/n_{.1}$ หรือ $f_{21}/n_{2.}$	$f_{22}/n_{.2}$ หรือ $f_{22}/n_{2.}$	$f_{23}/n_{.3}$ หรือ $f_{23}/n_{2.}$	1.00
รวม	1.00	1.00	1.00	

1.2 การคำนวณค่าความถี่ที่คาดหวัง (Expected Frequency)

การวิเคราะห์ข้อมูลเพื่อศึกษาความเป็นอิสระระหว่างตัวแปรนามบัญญัติโดยใช้สถิติไค-สแควร์ในสถิตินันพาราเมตริกทั่ว ๆ ไป นักวิจัยนำค่าความถี่สังเกตได้มาคำนวณหาความถี่ที่คาดหวัง แล้วนำไปคำนวณหาค่าสถิติไค-สแควร์เพื่อทดสอบสมมติฐานทางสถิติ การวิเคราะห์โมเดลล็อกลิเนียร์ใช้หลักการเดียวกัน มีข้อแตกต่างตรงที่การวิเคราะห์ล็อกลิเนียร์มีวิธีการคำนวณหาค่าความถี่ที่คาดหวังแตกต่างกันหลายวิธีตามข้อกำหนดในการสร้างโมเดลล็อกลิเนียร์

จากตัวอย่างตารางการณ์จร กรณี 2 ตัวแปร คือตัวแปร A และ B ในตารางขนาด i x j ที่กล่าวถึงในหัวข้อที่ 1.1 เมื่อ i = 2 และ j = 3 จะได้ตารางการณ์จร 2 แถว 3 สดมภ์ และมีจำนวนเซลล์ รวม 6 เซลล์ สามารถคำนวณค่าความถี่ที่คาดหวังในแต่ละเซลล์ได้ 4 ลักษณะ โดยมีวิธีการคำนวณดังนี้

เมื่อ p_{ij} = ความน่าจะเป็นในการจำแนกหน่วยตัวอย่างลงแถวที่ i สดมภ์ที่ j

n = จำนวนหน่วยตัวอย่าง หรือขนาดของกลุ่มตัวอย่างในการวิเคราะห์

F_{ij} = ความถี่ที่คาดหวังในแถวที่ i สดมภ์ที่ j

= np_{ij}

ลักษณะที่ 1 การคำนวณความถี่ที่คาดหวัง กรณีไม่มีอิทธิพลใด ๆ จากตัวแปร A และ B

ค่าความถี่ที่คาดหวังในแต่ละเซลล์ คือ จำนวนหน่วยตัวอย่างที่มีความน่าจะเป็นที่ถูกจำแนกลงในแถวที่ i สดมภ์ที่ j โดยไม่ได้รับอิทธิพลใดๆ จากตัวแปร A และ B นั่นคือ โอกาสที่หน่วยตัวอย่างถูกจำแนกลงเซลล์ ij มีค่าเท่ากัน ทำให้ความถี่ที่คาดหวังในเซลล์ ij มีค่าเท่า ๆ กันด้วย ผลการคำนวนค่าความถี่ที่คาดหวังแสดงได้ดังตารางที่ 3 ในกรณีนี้ตารางการณ์จรมี 6 เซลล์ และจำนวนหน่วยตัวอย่างมี n หน่วย ดังนั้น

$p_{ij} = 1/6$ และ $F_{ij} = n/6$

ตารางที่ 3 แสดงการแจกแจงความถี่ที่คาดหวังลงตารางการณ์จร 2 มิติ ขนาด 2 X 3 ลักษณะ 1

A \ B	1	2	3	รวม
1	n/6	n/6	n/6	$n_{1.}$
2	n/6	n/6	n/6	$n_{2.}$
รวม	$n_{.1}$	$n_{.2}$	$n_{.3}$	n

ลักษณะที่ 2 การคำนวณความถี่ที่คาดหวัง กรณีได้รับอิทธิพลหลักจากตัวแปร A

ค่าความถี่ที่คาดหวังในแต่ละเซลล์ คือ จำนวนหน่วยตัวอย่างที่มีความน่าจะเป็นที่ถูกจำแนกลงในแถว i สดมภ์ j โดยได้รับอิทธิพลหลักจากตัวแปร A เพียงตัวแปรเดียว ผลการคำนวณค่าความถี่ที่คาดหวังดังแสดงในตารางที่ 4 กรณีนี้ตารางการณ์จร มี 6 เซลล์ ตัวแปร A มี 2 ค่า เมื่อมีอิทธิพลจากตัวแปร A ผลรวมความถี่ที่คาดหวังในแถวที่ 1 และ 2 จึงแตกต่างกัน ทำให้ความถี่ที่คาดหวังในแต่ละสดมภ์ของแต่ละแถวมีค่าแตกต่างกันไปด้วย แต่ค่าความถี่ที่คาดหวังทั้ง 3 สดมภ์ในแต่ละแถวต้องเท่ากัน นั่นคือ โอกาสที่หน่วยตัวอย่างถูกจำแนกลงแต่ละสดมภ์ j ในแถว i มีค่าเท่า ๆ กัน โอกาสหรือค่าความน่าจะเป็นในที่นี้ คือ ความน่าจะเป็นมีเงื่อนไข (conditional probability=$p_{j/i}$) ที่กล่าวแล้วในตอนที่ 1.1 นั่นเอง เนื่องจากตารางการณ์จรในแต่ละแถวมี 3 สดมภ์ และจำนวนหน่วยตัวอย่างในแถว i หรือผลรวมความถี่แถว i มีค่าเท่ากับ $n_{i.}$ เมื่อความถี่สดมภ์ j ในแต่ละแถว i มีค่าเท่ากัน ค่า $p_{j/i} = 1/3$ จึงคำนวณหาความถี่ที่คาดหวังสดมภ์ j ในแต่ละแถว i ได้จากความน่าจะเป็นอย่างมีเงื่อนไข ดังนี้

$$F_{ij} = n_{i.}\, p_{j/i} = n_{i.} / 3$$

ตารางที่ 4 แสดงการแจกแจงความถี่ที่คาดหวังลงตารางการณ์จร 2 มิติ ขนาด 2 X 3 ลักษณะ 2

A \ B	1	2	3	รวม
1	$n_{1.} / 3$	$n_{1.} / 3$	$n_{1.} / 3$	$n_{1.}$
2	$n_{2.} / 3$	$n_{2.} / 3$	$n_{2.} / 3$	$n_{2.}$
รวม	$n_{.1}$	$n_{.2}$	$n_{.3}$	n

ลักษณะที่ 3 การคำนวณความถี่ที่คาดหวัง กรณีได้รับทั้งอิทธิพลหลักของตัวแปร A และอิทธิพลหลักของตัวแปร B

ค่าความถี่ที่คาดหวังในแต่ละเซลล์ คือ จำนวนหน่วยตัวอย่างที่มีความน่าจะเป็นที่ถูกจำแนกลงในแถว i สดมภ์ j โดยได้รับอิทธิพลทั้งอิทธิพลหลักของตัวแปร A และ B ดังนั้น ค่าความถี่ที่คาดหวังในแต่ละเซลล์จึงแตกต่างกันทั้งหมด ความน่าจะเป็นที่หน่วยตัวอย่างแต่ละหน่วยจะถูกจำแนกลงในแถว i สดมภ์ j (p_{ij}) มีค่าเท่ากับผลคูณของความน่าจะเป็นที่หน่วยตัวอย่างถูกจำแนกลงในแถว i ($p_{i.}$) กับความน่าจะเป็นที่หน่วยตัวอย่างถูกจำแนกลงในสดมภ์ j ($p_{.j}$) นั่นคือ

$p_{ij} = p_{i.}\, p_{.j}$ (ภายใต้ข้อตกลงเบื้องต้นว่าตัวแปร A และ B เป็นอิสระต่อกัน)

$= (n_{i.} / n)\,(n_{.j} / n)$

$F_{ij} = n\, p_{ij}$

$= n((n_{i.} / n)\,(n_{.j} / n))$

$= (n_{i.}\, n_{.j}) / n$

ผลการคำนวณค่าความถี่ที่คาดหวัง และความน่าจะเป็นแสดงในตารางที่ 5 จะเห็นได้ว่าค่าความถี่ที่คาดหวังในตารางการณ์จรมีลักษณะการคำนวณความถี่ที่คาดหวังซึ่งใช้ในการวิเคราะห์ความสัมพันธ์ หรือทดสอบความเป็นอิสระ (test of independence) ระหว่างตัวแปร 2 ตัวแปร โดยใช้สถิติไค-สแควร์นั่นเอง

ตารางที่ 5 แสดงการแจกแจงความถี่ที่คาดหวังลงตารางการณ์จร 2 มิติ ขนาด 2 X 3 ลักษณะ 3

A \ B	1	2	3	รวม
1	$n_{1.}\, n_{.1} / n$	$n_{1.}\, n_{.2} / n$	$n_{1.}\, n_{.3} / n$	$n_{1.}$
2	$n_{2.}\, n_{.1} / n$	$n_{2.}\, n_{.2} / n$	$n_{2.}\, n_{.3} / n$	$n_{2.}$
รวม	$n_{.1}$	$n_{.2}$	$n_{.3}$	n

ลักษณะที่ 4 การคำนวณความถี่ที่คาดหวัง กรณีได้รับทั้งอิทธิพลหลักของตัวแปร A อิทธิพลหลักของตัวแปร B และปฏิสัมพันธ์ระหว่าง A กับ B

ค่าความถี่ที่คาดหวังในแต่ละเซลล์ คือ จำนวนหน่วยตัวอย่างที่มีความน่าจะเป็นที่ถูกจำแนกลงในแถว i สดมภ์ j โดยได้รับอิทธิพลหลักของตัวแปร A อิทธิพลหลักของตัวแปร B และปฏิสัมพันธ์ระหว่างตัวแปร A และ B ค่าความถี่ที่คาดหวังในแต่ละเซลล์มีค่าเท่ากับความถี่สังเกตได้ ดังแสดงในตารางที่ 6

ตารางที่ 6 แสดงการแจกแจงความถี่ที่คาดหวังลงตารางการณ์จร 2 มิติ ขนาด 2 X 3 ลักษณะ 4

A \ B	1	2	3	รวม
1	f_{11}	f_{12}	f_{13}	$n_{1.}$
2	f_{21}	f_{22}	f_{23}	$n_{2.}$
รวม	$n_{.1}$	$n_{.2}$	$n_{.3}$	n

ลักษณะการคำนวณค่าความถี่ที่คาดหวังทั้ง 4 ลักษณะที่กล่าวแล้วข้างต้น สะท้อนให้เห็นถึงอิทธิพลหลักและปฏิสัมพันธ์ จากตัวแปร A และ B ซึ่งมีลักษณะคล้ายคลึงกับโมเดลในการวิเคราะห์ความแปรปรวนแบบสองทาง (two way ANOVA) ประเด็นที่แตกต่างกันคือ การวิเคราะห์ล็อกลิเนียร์ นักวิจัยศึกษาค่าความถี่ที่คาดหวัง ในขณะที่การวิเคราะห์ความแปรปรวน นักวิจัยศึกษาค่าเฉลี่ยของตัวแปรตาม เมื่อเปรียบเทียบลักษณะการคำนวณความถี่ที่คาดหวัง 4 ลักษณะนี้กับลักษณะการหาค่าเฉลี่ยในการวิเคราะห์ความแปรปรวนทั้ง 4 ลักษณะจะเทียบเคียงกันได้เป็น 4 สมการ เมื่อกำหนดให้

Y_{ij} = คะแนนเฉลี่ยเซลล์ ij
F_{ij} = ความถี่ที่คาดหวังในเซลล์ ij
$\bar{Y}..$ = ค่าเฉลี่ยรวม = $\sum_{i=1}^{2}\sum_{j=1}^{2}\frac{Y_{ij}}{n}$
τ = ค่าเฉลี่ยรวมของความถี่ที่คาดหวัง
a_i = อิทธิพลหลักจากตัวแปร A
τ_i^a = อิทธิพลหลักจากตัวแปร A
b_j = อิทธิพลหลักจากตัวแปร B
τ_j^b = อิทธิพลหลักจากตัวแปร B
ab_{ij} = อิทธิพลปฏิสัมพันธ์ระหว่างตัวแปร A กับ B
τ_{ij}^{ab} = อิทธิพลปฏิสัมพันธ์ระหว่างตัวแปร A กับ B

นั่นคือ

ลักษณะสมการที่	การวิเคราะห์ความแปรปรวน	การวิเคราะห์ล็อกลิเนียร์	
1	$Y_{ij} = \bar{Y}..$	$F_{ij} = \tau$	$= n/6 = np_{ij}$
2	$Y_{ij} = \bar{Y}.. + a_i$	$F_{ij} = \tau\tau_i^a$	$= n_{i.}/3 = np_{i.}$
3	$Y_{ij} = \bar{Y}.. + a_i + b_j$	$F_{ij} = \tau\tau_i^a\tau_j^b$	$= n_{i.}\, n_{.j}/n = np_{i.}p_{.j}$
4	$Y_{ij} = \bar{Y}.. + a_i + b_j + ab_{ij}$	$F_{ij} = \tau\tau_i^a\tau_j^b\tau_{ij}^{ab}$	$= f_{ij}$

จะเห็นได้ว่าสมการในการวิเคราะห์ความแปรปรวน เป็นสมการมุ่งอธิบายค่าเฉลี่ยของตัวแปรตามในแต่ละเซลล์โดยใช้สมการผลบวกของอิทธิพลจากตัวแปรต้น แต่สมการในการวิเคราะห์ตารางการณ์จรเป็นสมการมุ่งอธิบายค่าความถี่ที่คาดหวังในแต่ละเซลล์โดยใช้ผลคูณของอิทธิพลจากตัวแปร (Kennedy, 1983) ในการวิเคราะห์ความแปรปรวน นักวิจัยนำสมการลักษณะที่ 4 ที่เป็นการศึกษาทั้งอิทธิพลหลักและปฏิสัมพันธ์ของตัวแปร A และ B ใช้ในการวิเคราะห์ข้อมูลโดยมิได้นำสมการ 3 สมการแรกไปวิเคราะห์ด้วย แต่ตามหลักการวิเคราะห์สาเหตุ (causal analysis) Kennedy (1983) กล่าวว่า สมการในการวิเคราะห์ 3 สมการแรก **เป็นสมการที่ประหยัด และเข้มงวดมากที่สุด** (most restriction) **ส่วนสมการสุดท้าย** มีความประหยัดและเข้มงวดน้อยที่สุด (least restriction) จึงควรมีการนำสมการทั้ง 4 ลักษณะไปตรวจสอบก่อนว่าข้อมูลมีลักษณะสอดคล้องกับสมการในลักษณะใด แล้วจึงจะสามารถนำ

สมการที่ตรวจสอบได้นี้ไปใช้ต่อไป กระบวนการนำสมการดังกล่าวไปใช้และการตรวจสอบความเหมาะสม คือ กระบวนการสร้างโมเดลและตรวจสอบโมเดล

2. การสร้างโมเดลล็อกลิเนียร์ (Log-Linear Modeling)

โมเดลล็อกลิเนียร์ คือ สมการคณิตศาสตร์ที่นักวิจัยสร้างขึ้น เพื่อประมาณค่าความถี่ที่คาดหวังแต่ละเซลล์ของตารางการณ์จรในรูปผลคูณของเทอมอิทธิพลของตัวแปร เช่น สมการที่กล่าวแล้ว หน้า 18 จะเห็นว่า โมเดลล็อกลิเนียร์มีลักษณะคล้ายโมเดลการวิเคราะห์ความแปรปรวน ต่างกันที่โมเดลล็อกลิเนียร์มีลักษณะโมเดลทุกรูปแบบ โดยมิได้ขจัดรูปแบบอื่นทิ้ง ดังนั้น โมเดลล็อกลิเนียร์ กรณีตัวแปร 2 ตัวแปรจึงมีโมเดลได้ 4 แบบ ในขณะที่โมเดลการวิเคราะห์ความแปรปรวนมีได้แบบเดียวไม่ว่าจะมีจำนวนตัวแปรเท่าใดก็ตาม ประเด็นที่แตกต่างอีกอย่างคือ โมเดลการวิเคราะห์ความแปรปรวนเป็นโมเดลแบบบวก ในขณะที่โมเดลล็อกลิเนียร์เป็นแบบคูณ

ลักษณะโมเดลล็อกลิเนียร์ที่เป็นแบบคูณ แสดงถึงความสัมพันธ์ที่ไม่ใช่เส้นตรง Kennedy (1983) และ Steven (1996) กล่าวว่าโมเดลแบบคูณนี้อาจจะแปลงให้เป็นโมเดลเชิงเส้นโดยการคำนวณหาลอการิทึม โดยทั่วไปนิยมใช้ลอการิทึมธรรมชาติ (natural logarithm) ซึ่งอยู่ในรูปฐาน e และ e เป็นค่าคงที่ มีค่าประมาณ 2.718 ดังนั้น โมเดลล็อกลิเนียร์ที่ได้จากการคำนวณความถี่ที่คาดหวัง 4 แบบ ซึ่งเสนอไว้ในหัวข้อที่ 1.2 ทำให้เป็นโมเดลเชิงเส้นได้ดังนี้

กำหนดให้ $\lambda = \ln \tau$, $\lambda_i^a = \ln \tau_i^a$, $\lambda_j^b = \ln \tau_j^b$, $\lambda_{ij}^{ab} = \ln \tau_{ij}^{ab}$

จากสมการที่ 1	$F_{ij} = \tau$	ดังนั้น	$\ln F_{ij} = \lambda$
จากสมการที่ 2	$F_{ij} = \tau\tau_i^a$		$\ln F_{ij} = \lambda + \lambda_i^a$
จากสมการที่ 3	$F_{ij} = \tau\tau_i^a\tau_j^b$		$\ln F_{ij} = \lambda + \lambda_i^a + \lambda_j^b$
จากสมการที่ 4	$F_{ij} = \tau\tau_i^a\tau_j^b\tau_{ij}^{ab}$		$\ln F_{ij} = \lambda + \lambda_i^a + \lambda_j^b + \lambda_{ij}^{ab}$

โมเดลล็อกลิเนียร์ แบ่งออกเป็นหลายประเภทตามลักษณะของการคำนวณค่าความถี่ที่คาดหวัง ในกรณีที่มีตัวแปร 2 ตัวแปรการคำนวณค่าความถี่ที่คาดหวังมี 4 ลักษณะ ได้สมการอธิบายค่าความถี่ที่คาดหวัง 4 รูปแบบ นั้น เป็นที่มาของโมเดลล็อกลิเนียร์ทั่ว ๆ ไป (general log-linear model) โมเดลทั้ง 4 โมเดล คือ โมเดลความน่าจะเป็นเท่า (mutual equiprobability model) โมเดลความน่าจะเป็นเท่ากันแบบมีเงื่อนไข (conditional equiprobability model) โมเดล

หอสมุดกลาง สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

อิสระต่อกัน (mutual independence model) และโมเดลอิ่มตัว (saturated model) ดังรายละเอียดต่อไปนี้

2.1 โมเดลความน่าจะเป็นเท่า (Mutual Equiprobability Model) หรือโมเดลศูนย์ (Null Model)

โมเดลนี้ทำนายความถี่ที่คาดหวังโดยไม่ขึ้นกับอิทธิพลใดๆ ของตัวแปรเลย จากลักษณะการทำนายความถี่ที่คาดหวังนี้ชี้ให้เห็นว่าหน่วยตัวอย่างมีคุณลักษณะ (character) ที่เหมือนๆ กัน เป็นลักษณะที่คงที่ ในที่นี้ คือ ค่าเฉลี่ยรวมของความถี่ที่คาดหวังหรือ τ นั่นเอง และการคำนวณความถี่ที่คาดหวังใช้จำนวนหน่วยตัวอย่างทั้งหมดเป็นฐานในการคำนวณจึงใช้สัญญลักษณ์ [n] บอกถึงผลรวมความถี่ที่ใช้เป็นฐานในการคำนวณ ดังนั้นโมเดลล็อกลิเนียร์มีลักษณะดังนี้

$$F_{ij} = \tau = [n]$$

2.2 โมเดลความน่าจะเป็นเท่ากันแบบมีเงื่อนไข (Conditional Equiprobability Model)

โมเดลนี้ทำนายความถี่ที่คาดหวังเมื่อได้รับอิทธิพลหลักของ A ทำให้ความถี่ที่คาดหวังในแต่ละเซลล์ซึ่งควรจะเท่ากันดังโมเดลที่ 1 กลับมีค่าแตกต่างกัน ยกเว้นความถี่ที่คาดหวังทุกสดมภ์ในแต่ละแถวของ A มีค่าเท่ากัน และการคำนวณความถี่ที่คาดหวังใช้จำนวนหน่วยตัวอย่างในแถวของ A เป็นฐาน จึงใช้สัญญลักษณ์ [A] แทนฐานข้อมูลที่ใช้ในการคำนวณความถี่ที่คาดหวัง ดังนั้น โมเดลล็อกลิเนียร์มีลักษณะ ดังนี้

$$F_{ij} = \tau\tau_i^a = [A]$$

2.3 โมเดลอิสระต่อกัน (Mutual Independence Model)

โมเดลนี้ทำนายความถี่ที่คาดหวังที่ได้รับอิทธิพลหลักของ A อิทธิพลหลักของ B การคำนวณความถี่ที่คาดหวังใช้ผลรวมจำนวนตัวอย่างในแถว i สดมภ์ j เป็นฐาน จึงใช้สัญญลักษณ์ [A][B] แทนฐานข้อมูลที่ใช้ในการคำนวณความถี่ที่คาดหวัง ดังนั้น โมเดลล็อกลิเนียร์มีลักษณะ ดังนี้

$$F_{ij} = \tau\tau_i^a\tau_j^b = [A][B]$$

2.4 โมเดลอิ่มตัว (Saturated model)

โมเดลนี้ทำนายความถี่ที่คาดหวังที่ได้รับอิทธิพลหลักของ A อิทธิพลหลักของ B และปฏิสัมพันธ์ระหว่าง A กับ B ความถี่ที่คาดหวังเท่ากับความถี่ที่สังเกตได้เซลล์ ij นั่นคือ ความถี่

ที่คาดหวังคำนวณโดยใช้จำนวนตัวอย่างในเซลล์ ij เป็นฐาน จึงใช้สัญญลักษณ์ [AB] แทนฐานข้อมูลที่ใช้ในการคำนวณความถี่ที่คาดหวัง ดังนั้นโมเดลล็อกลิเนียร์มีลักษณะดังนี้

$$F_{ij} = \tau\tau_i^a\tau_j^b\tau_{ij}^{ab} = [AB]$$

กรณีตัวแปร 2 ตัวแปรมีโมเดลล็อกลิเนียร์ทำนายความถี่ที่คาดหวัง 4 โมเดล แต่ละโมเดลมีจำนวนเทอมอิทธิพลแตกต่างกัน และการประมาณค่าเทอมอิทธิพลได้จากความถี่ที่คาดหวังที่ประมาณได้ของแต่ละโมเดล ดังรายละเอียดหัวข้อต่อไปนี้

3. การประมาณค่าพารามิเตอร์ (Parameter Estimation)

การวิเคราะห์ล็อกลิเนียร์มีหลักการประมาณค่าพารามิเตอร์ คือ เทอมอิทธิพลต่าง ๆ ในแต่ละโมเดลเหมือนกับการประมาณค่าพารามิเตอร์อิทธิพลของตัวแปรในโมเดลการวิเคราะห์ความแปรปรวน แต่เนื่องจากโมเดลการวิเคราะห์ความแปรปรวนมีเพียง 1 โมเดล และการประมาณค่าเทอมอิทธิพลต่าง ๆ มีดังต่อไปนี้

$$Y_{ij} = \mu + a_i + b_j + ab_{ij}$$

พารามิเตอร์ที่ต้องประมาณค่าคือ μ, a_i, b_j, ab_{ij} เมื่อกำหนดให้

$\bar{Y}_{..}$ = ค่าเฉลี่ยทั้งหมดของ Y_{ij} หรือ grand mean = $\sum_{i=1}^{2}\sum_{j=1}^{2}\frac{Y_{ij}}{n}$

$\bar{Y}_{i.}$ = ค่าเฉลี่ยของ Y_{ij} ในแต่ละระดับของ A หรือ group mean ของ A = $\sum_{j=1}^{2}\frac{Y_{ij}}{n_{i.}}$

$\bar{Y}_{.j}$ = ค่าเฉลี่ยของ Y_{ij} ในแต่ละระดับของ B หรือ group mean ของ B = $\sum_{i=1}^{2}\frac{Y_{ij}}{n_{.j}}$

ประมาณค่าพารามิเตอร์ได้ดังนี้

$\mu = \bar{Y}_{..}$

$a_i = \bar{Y}_{i.} - \bar{Y}_{..}$

$b_j = \bar{Y}_{.j} - \bar{Y}_{..}$

$ab_{ij} = Y_{ij} - a_i - b_j + \bar{Y}_{..}$

การวิเคราะห์ล็อกลิเนียร์มี 4 โมเดลแต่ละโมเดลเทียบได้กับโมเดลการวิเคราะห์ความแปรปรวนดังที่กล่าวแล้วหน้า 18 พารามิเตอร์ที่ต้องประมาณในโมเดลล็อกลิเนียร์แต่ละโมเดล คือ เทอมอิทธิพล τ แต่ละเทอม และเมื่อกำหนดให้

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

$G_{..}$ แทน มัชฌิมเรขาคณิตทั้งหมดของความถี่ที่คาดหวัง (grand geometric mean)

$= (\prod F_{ij})^{1/ij}$

$G_{i.}^{a}$ แทน มัชฌิมเรขาคณิตของความถี่ที่คาดหวังของกลุ่มหรือในแต่ละระดับของ A

(group geometric mean) $= (\prod F_{ij})^{1/j}$

$G_{.j}^{b}$ แทน มัชฌิมเรขาคณิตของความถี่ที่คาดหวังของกลุ่มหรือในแต่ละระดับของ B

(group geometric mean) $= (\prod F_{ij})^{1/i}$

โดยที่ $\prod$ หมายถึง ผลคูณ

นั่นคือ

<u>โมเดลล็อกลิเนียร์แบบคูณ</u>	<u>การประมาณค่าอิทธิพล</u>
$F_{ij} = \tau$	$\tau = G_{..}$
$F_{ij} = \tau\tau_i^a$	$\tau = G_{..}$, $\tau_i^a = G_i^a / G_{..}$
$F_{ij} = \tau\tau_i^a\tau_j^b$	$\tau = G_{..}$, $\tau_i^a = G_i^a / G_{..}$, $\tau_j^b = G_j^b / G_{..}$
$F_{ij} = \tau\tau_i^a\tau_j^b\tau_{ij}^{ab}$	$\tau = G_{..}$, $\tau_i^a = G_i^a / G_{..}$, $\tau_j^b = G_j^b / G_{..}$, $\tau_{ij}^{ab} = f_{ij} / \tau\tau_i^a\tau_j^b$

<u>โมเดลล็อกลิเนียร์แบบบวก</u>	<u>การประมาณค่าอิทธิพล</u>
$\ln F_{ij} = \lambda$	$\lambda = \ln G_{..}$
$\ln F_{ij} = \lambda+\lambda_i^a$	$\lambda = \ln G_{..}$, $\lambda_i^a = \ln G_i^a - \ln G_{..}$
$\ln F_{ij} = \lambda+\lambda_i^a+\lambda_j^b$	$\lambda = \ln G_{..}$, $\lambda_i^a = \ln G_i^a - \ln G_{..}$, $\lambda_j^b = \ln G_j^b - \ln G_{..}$
$\ln F_{ij} = \lambda+\lambda_i^a+\lambda_j^b+\lambda_{ij}^{ab}$	$\lambda = \ln G_{..}$, $\lambda_i^a = \ln G_i^a - \ln G_{..}$, $\lambda_j^b = \ln G_j^b - \ln G_{..}$, $\lambda_{ij}^{ab} = \ln f_{ij} - \ln G_i^a - \ln G_j^b + \ln G_{..}$

โดยมีข้อตกลงสำหรับการประมาณค่าเทอมอิทธิพลจากโมเดลแบบคูณ คือ

$$\prod\tau_i^a = 1 , \quad \prod\tau_j^b = 1 , \quad \prod\tau_{ij}^{ab} = 1$$

และมีข้อตกลงสำหรับการประมาณค่าเทอมอิทธิพลจากโมเดลแบบบวก คือ

$$\sum_{i=1}^{2}\lambda_i^a = 0 , \quad \sum_{j=1}^{3}\lambda_j^b = 0 , \quad \sum_{i=1}^{2}\sum_{j=1}^{3}\lambda_{ij}^{ab} = 0$$

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

การประมาณค่าพารามิเตอร์เทอมอิทธิพลของโมเดลทั้ง 4 ไม่ว่าจะเป็นแบบคูณหรือแบบบวก มีข้อสังเกต คือ เทอมอิทธิพลคงที่แต่ละโมเดลมีค่าต่างกัน เทอมอิทธิพลแต่ละระดับของ A โมเดลที่ 2 และ 3 มีค่าเท่ากัน แต่จะต่างกับโมเดลที่ 4 และเทอมอิทธิพลแต่ละระดับของ B โมเดลที่ 3 และ 4 มีค่าต่างกัน เนื่องจากโมเดลที่ 2 ได้รับอิทธิพลหลักของ A ทำให้อิทธิพลคงที่มีค่าเปลี่ยนแปลง และโมเดลที่ 3 ได้รับอิทธิพลหลักของ B เพิ่มเข้ามาแต่ไม่เกี่ยวข้องกับอิทธิพลของ A ทำให้ค่าอิทธิพลของ A ไม่เปลี่ยนแปลง สำหรับโมเดลที่ 4 ได้รับปฏิสัมพันธ์ระหว่าง A และ B ทำให้ค่าอิทธิพลของ A และ B เปลี่ยนแปลงไป

กรณีตัวแปร 2 ตัวแปร ค่าความถี่ที่คาดหวังและค่าอิทธิพลต่าง ๆ สามารถคำนวณได้ด้วยเครื่องคิดเลขอัตโนมัติแต่ถ้ามีตัวแปรตั้งแต่ 3 ตัวแปรขึ้นไป การคำนวณค่าความถี่ที่คาดหวังและค่าอิทธิพลต่างๆ ไม่สามารถจะคำนวณได้ นักสถิติหลายท่าน ได้แก่ R.A. Fisher , Deming และ Stephan, Birch, Fienberg, Goodman และ Haberman ได้พัฒนาการประมาณค่าด้วยวิธีความเป็นไปได้สูงสุด (maximum likelihood , ML) ด้วยหลักการคำนวณทวนซ้ำ (iterative proportional fitting algorithm) และพัฒนาโปรแกรมคอมพิวเตอร์ช่วยอำนวยความสะดวกในการประมาณค่าดังกล่าว (Kennedy, 1983; Hagenaars, 1993)

จากที่กล่าวมาจะเห็นว่าโมเดลล็อกลิเนียร์ กรณีตัวแปร 2 ตัวแปร มีโมเดลที่มีลักษณะต่างกัน 4 ลักษณะจึงจำเป็นที่จะต้องคัดเลือกโมเดลเพียงโมเดลเดียวที่มีความกลมกลืนกับข้อมูลและโมเดลที่เลือกเป็นโมเดลแบบประหยัด (parsimonious model) ดังหัวข้อต่อไป

4. **การคัดเลือกโมเดลล็อกลิเนียร์** (Log-Linear Models Selection)

สำหรับการศึกษาตัวแปร 2 ตัวแปร นักวิจัยสร้างโมเดลล็อกลิเนียร์มีลักษณะลดหลั่นกันได้ 4 โมเดล แต่ละโมเดลสะท้อนถึงอิทธิพลของตัวแปรต่างกัน ดังตารางที่ 7 ที่แสดงลักษณะโมเดลและความหมายของโมเดล 4 โมเดล ผู้วิจัยจะต้องคัดเลือกโมเดลล็อกลิเนียร์ 4 โมเดลนี้ให้เหลือเพียงโมเดลเดียวที่เหมาะสมกับการอธิบายความถี่ที่คาดหวังโดยการทดสอบความกลมกลืน (test of goodness of fit) ระหว่างความถี่ที่คาดหวัง (F_{ij}) กับข้อมูลเชิงประจักษ์หรือความถี่ที่สังเกตได้ (f_{ij}) โมเดลล็อกลิเนียร์ที่มีความกลมกลืนกับข้อมูลเชิงประจักษ์และง่ายที่สุดต่อการอธิบายการทำนายความถี่ที่คาดหวังซึ่งสามารถพิจารณาความสัมพันธ์หรือความแตกต่างของตัวแปรได้อย่างมีนัยสำคัญทางสถิติ เรียกว่าโมเดลประหยัด (parsimonious model) (Kennedy, 1983; Steven, 1996) การคัดเลือกโมเดลล็อกลิเนียร์ที่เป็นโมเดลประหยัดมีขั้นตอนการทดสอบความกลมกลืนของโมเดล 2 ขั้นตอน Kennedy (1983) ได้อธิบายไว้ดังนี้

ขั้นตอนแรก ทดสอบความกลมกลืนระหว่างความถี่ที่คาดหวัง (F) ของโมเดลแต่ละลักษณะกับความถี่ที่สังเกตได้ (f) จากสมมติฐานหลัก คือ H_0 : โมเดลมีความกลมกลืนกับข้อมูลเชิงประจักษ์ การทดสอบใช้ค่าสถิติไค-สแควร์ (χ^2) ที่พัฒนาโดย Pearson หรือ อัตราส่วนไลค์ลี่ฮูดไค-สแควร์ (likelihood ratio chi-square) สัญญลักษณ์ที่ใช้ คือ L^2 พัฒนาโดย R.A. Fisher ดังสูตรการคำนวณเฉพาะกรณีการวิเคราะห์ข้อมูลของ 2 ตัวแปร ต่อไปนี้

ไค-สแควร์ $$\chi^2 = \sum\sum \frac{(f_{ij} - F_{ij})^2}{F_{ij}}$$

อัตราส่วนไลค์ลี่ฮูดไค-สแควร์ $$L^2 = 2 \sum\sum (f_{ij}) \ln\left(\frac{f_{ij}}{F_{ij}}\right)$$

ค่าสถิติทั้ง 2 ค่าอาจมีค่าต่างกันแต่ต่างกันไม่มากนัก โดยเฉพาะถ้าจำนวนหน่วยตัวอย่างมีขนาดใหญ่ สถิติทั้ง 2 มีการแจกแจงแบบไค-สแควร์ และมีค่าไกล้เคียงกันมากทำให้ผลสรุปการทดสอบสมมติฐานเหมือนกัน แต่สถิติอัตราส่วนไลค์ลิฮูดไค-สแควร์จะเป็นที่นิยมใช้มากกว่า เพราะสถิติอัตราส่วนไลค์ลิฮูดไค-สแควร์สามารถแบ่งเป็นหลายค่า แต่ละค่าแสดงถึงอิทธิพลของแต่ละเทอมในโมเดล (Goodman, 1971)

ตารางที่ 7 แสดงลักษณะและความหมายของโมเดลล็อกลิเนียร์ 4 โมเดล สำหรับตารางขนาด 2 มิติ

โมเดลที่	ลักษณะของโมเดลแบบคูณ	ลักษณะของโมเดลตามฐานข้อมูล	ความหมาย
1	$F_{ij} = \tau$	[n]	ความถี่ที่คาดหวังทำนายได้ด้วยอิทธิพลคงที่เท่านั้น และทำให้ความถี่ที่คาดหวังปรากฏในทุกเซลล์ด้วยโอกาสเท่า ๆ กัน สรุปได้ว่าตัวแปร A และ B ไม่มีความสัมพันธ์กัน
2	$F_{ij} = \tau\tau_i^a$	[A]	ความถี่ที่คาดหวังทำนายได้ด้วยอิทธิพลคงที่และอิทธิพลหลักของ A เท่านั้น จากอิทธิพลของ A ทำให้ความถี่ที่คาดหวังในแต่ละกลุ่มของ B มีค่าเท่ากันในแต่ละกลุ่มของ A สรุปได้ว่าตัวแปร A และ B ไม่มีความสัมพันธ์กัน

ตารางที่ 7 (ต่อ)

โมเดลที่	ลักษณะของโมเดลแบบคูณ	ลักษณะของโมเดลตามฐานข้อมูล	ความหมาย
3	$F_{ij} = \tau\tau_i^a\tau_j^b$	[A][B]	ความถี่ที่คาดหวังทำนายได้ด้วยอิทธิพลคงที่ อิทธิพลหลักของ A และ B เท่านั้น ไม่มีปฏิสัมพันธ์ของ A และ B อิทธิพลหลักทั้งสองทำให้ความถี่ที่คาดหวังในแต่ละกลุ่มของ A และ B มีความแตกต่างกัน แต่จะมีค่าใกล้เคียงกับค่าความถี่ที่สังเกตได้ สรุปได้ว่าตัวแปร A และ B ไม่มีความสัมพันธ์กัน
4	$F_{ij} = \tau\tau_i^a\tau_j^b\tau_{ij}^{ab}$	[AB]	ความถี่ที่คาดหวังทำนายได้ด้วยอิทธิพลคงที่ อิทธิพลหลักของ A อิทธิพลหลักของ B และปฏิสัมพันธ์ A และ B ทำให้ความถี่ที่คาดหวังทุกเซลล์มีค่าเท่ากับความถี่ที่สังเกตได้ สรุปได้ว่า ตัวแปร A และ B มีความสัมพันธ์กัน

จากค่าสถิติ L^2 ที่คำนวณได้ของแต่ละโมเดลเปรียบเทียบกับค่าไค-สแควร์แต่ละองศาอิสระ (degrees of freedom) จากตารางสถิติ คือ $\chi^2_{\alpha,df}$ แล้วพิจารณาคัดโมเดลที่ผลการทดสอบสมมติฐานไม่ยอมรับสมมติฐานหลัก (reject H_0) ออกจากการพิจารณา สำหรับองศาอิสระ (degrees of freedom) หรือ df ของโมเดลที่นำมาทดสอบ Kennedy (1983) เสนอว่า คำนวณได้จากสูตร

องศาอิสระของโมเดลทดสอบ = ผลรวมของจำนวนเทอมอิทธิพลในแต่ละค่าของตัวแปรจากโมเดลอิ่มตัวที่ประมาณค่าได้อย่างอิสระ ลบออกด้วยผลรวมของจำนวนเทอมอิทธิพลในแต่ละค่าของตัวแปรจากโมเดลที่ต้องการทดสอบและประมาณค่าได้อย่างอิสระ

ตัวอย่างการคำนวณองศาอิสระของโมเดลที่ 1 และ 2 ของตารางที่ 7 กรณีตัวแปร 2 ตัวแปร คือ A และ B โดยที่ A มี 2 ค่า และ B มี 3 ค่า นั้น การพิจารณาจำนวนเทอมอิทธิพลที่ประมาณได้อย่างอิสระทำได้ดังนี้

τ คือ เทอมของอิทธิพลคงที่มีเพียงค่าเดียวที่ประมาณค่าได้ ดังนั้น จำนวนเทอมที่ประมาณได้อย่างอิสระมีค่าเท่ากับ 1

τ_i^a คือ เทอมของอิทธิพลหลักของ A ซึ่งมีเทอมอิทธิพลในแต่ละค่าของ A 2 เทอม คือ τ^a_1 , τ^a_2 โดยมีข้อตกลงว่า $\tau^a_1 \tau^a_2 = 1$ หรือ $\prod\tau^a_i = 1$ ($\prod$ คือ ผลคูณ) นั่นคือ จะประมาณค่า τ^a_1 หรือ τ^a_2 เทอมใดเทอมหนึ่งอย่างอิสระได้ 1 เทอม ส่วนเทอมที่เหลือประมาณค่าได้ตามข้อตกลง ฉะนั้น อิทธิพลของ A แต่ละค่าสามารถประมาณค่าได้อย่างอิสระมีจำนวน 1 เทอม

τ_j^b คือ เทอมของอิทธิพลหลักของ B ซึ่งมีเทอมอิทธิพลในแต่ละค่าของ B 3 เทอม คือ τ^b_1 , τ^b_2 , τ^b_3 โดยมีข้อตกลงว่า $\tau^b_1 \tau^b_2 \tau^b_3 = 1$ หรือ $\prod\tau^b_j = 1$ ($\prod$ คือ ผลคูณ) นั่นคือ จะประมาณค่าอิทธิพลอย่างอิสระได้ 2 เทอม ส่วนเทอมที่เหลือประมาณค่าได้ตามข้อตกลง ฉะนั้น อิทธิพลของ B แต่ละค่าสามารถประมาณค่าได้อย่างอิสระมีจำนวน 2 เทอม

τ_{ij}^{ab} คือ เทอมของปฏิสัมพันธ์ AB มี 6 เทอม คือ τ_{11}^{ab}, τ_{12}^{ab}, τ_{13}^{ab}, τ_{21}^{ab}, τ_{22}^{ab}, τ_{23}^{ab} โดยมีข้อตกลง $\prod\tau_{ij}^{ab} = \prod\tau_{ij}^{ab} = 1$ ($\prod$ คือ ผลคูณ) และเนื่องจากประมาณค่าเทอมอิทธิพลแต่ละค่าของ A อย่างอิสระได้ 1 เทอม และประมาณค่าเทอมอิทธิพลของ B แต่ละค่าอย่างอิสระได้ 2 เทอม ทำให้จำนวนเทอมอิทธิพลทั้ง 6 เทอมสามารถประมาณค่าอย่างอิสระได้ 1 X 2 = 2 เทอม

จากรายละเอียดข้างต้นนี้สามารถคำนวณผลรวมของจำนวนเทอมอิทธิพลในแต่ละค่าของตัวแปรในโมเดลอิ่มตัว ($F_{ij} = \tau\tau_i^a\tau_j^b\tau_{ij}^{ab}$) ที่ประมาณค่าอย่างอิสระได้เท่ากับ 1+1+2+2 = 6 และคำนวณผลรวมของจำนวนเทอมอิทธิพลในแต่ละค่าของตัวแปรในโมเดลลักษณะที่ 1 ($F_{ij} = \tau$) ที่ประมาณค่าอย่างอิสระได้เท่ากับ 1 ดังนั้น องศาอิสระของโมเดลลักษณะที่ 1 มีค่าเท่ากับ 6-1= 5 ส่วนผลรวมของจำนวนเทอมอิทธิพลในแต่ละค่าของตัวแปรในโมเดลลักษณะที่ 2 ($F_{ij} = \tau\tau_i^a$) ที่ประมาณค่าอย่างอิสระได้เท่ากับ 1+1= 2 ดังนั้น องศาอิสระของโมเดลลักษณะที่ 2 มีค่าเท่ากับ 6 - 2 = 4

ขั้นตอนที่สอง ทดสอบนัยสำคัญของเทอมอิทธิพลที่เพิ่มเข้าไปในโมเดลล็อกลิเนียร์ที่ผ่านการทดสอบในขั้นตอนแรกทีละเทอมว่ามีอิทธิพลเพียงพอที่จะทำให้โมเดลล็อกลิเนียร์มีความกลมกลืนกับข้อมูลเชิงประจักษ์ โดยตั้งสมมติฐานหลักว่า H_0 : อิทธิพลตัวใหม่ที่เพิ่มเข้าไปมีค่าเท่ากับศูนย์ การทดสอบทำได้ด้วยการคำนวณผลต่างของค่าสถิติ L^2 (ใช้สัญญลักษณ์ ΔL^2)

ของโมเดลล็อกลิเนียร์ที่ทดสอบเป็นคู่ ๆ โดยโมเดลล็อกลิเนียร์แต่ละคู่ต้องมีลักษณะลดหลั่นเรียงกันไป คือ โมเดลล็อกลิเนียร์ที่มีความกลมกลืนกับข้อมูลแล้วนำมาทดสอบนัยสำคัญเทอมอิทธิพลเทอมสุดท้ายจะสามารถทดสอบคู่กับโมเดลอื่นที่มีเทอมอิทธิพลน้อยกว่าอยู่ 1 เทอม แต่โมเดลล็อกลิเนียร์ที่มีเทอมอิทธิพลน้อยกว่านั้นต้องมีเทอมอิทธิพลเหมือนกับโมเดลล็อกลิเนียร์ที่ทดสอบนัยสำคัญด้วย เมื่อคำนวณค่า ΔL^2 แล้วนำมาเปรียบเทียบกับค่าไค-สแควร์จากตารางสถิติ คือ $\chi^2_{\alpha,df}$ ซึ่ง df คำนวณได้จากผลต่างขององศาอิสระของโมเดลล็อกลิเนียร์คู่ที่ทดสอบคู่นั้น ๆ แล้วพิจารณาเลือกผลการทดสอบที่ปฏิเสธสมมติฐาน (reject H_0) ซึ่งแสดงว่าเทอมอิทธิพลนั้น ๆ มีอิทธิพลเพียงพอที่ทำให้โมเดลมีความกลมกลืนกับข้อมูลเชิงประจักษ์ จึงสมควรเลือกโมเดลนี้มากกว่าอีกโมเดลที่เป็นคู่ทดสอบ

ยกตัวอย่างเช่น การเลือกโมเดลล็อกลิเนียร์ของตัวแปร 2 ตัวแปร คือ A และ B โดยที่ A มี 2 ค่า และ B มี 2 ค่า กำหนดโมเดลต่างกันได้ 4 โมเดล และผลการคำนวณค่าต่าง ๆ ดังตารางที่ 8

ตารางที่ 8 ค่าสถิติอัตราส่วนไลค์ลิฮูดไค-สแควร์ และผลต่างของค่าสถิติอัตราส่วนไลค์ลิฮูดไค-สแควร์ของโมเดลล็อกลิเนียร์ กรณี 2 ตัวแปร

โมเดลที่	โมเดลแบบคูณ	H_0: Model กลมกลืนกับข้อมูล			H_0: เทอมอิทธิพลมีค่าเท่ากับศูนย์			
		L^2	DF	P	โมเดล	ΔL^2	DF	P
1	τ	25.68	3	.001				
2	$\tau\tau^a$	21.65	2	.001	1, 2	4.03	1	.050
3	$\tau\tau^a\tau^b$	5.19	1	.025	2, 3	16.46	1	.001
4	$\tau\tau^a\tau^b\tau^{ab}$	0.00	0	1.000	3, 4	5.19	1	.025

หมายเหตุ α=.01

จากตารางที่ 8 โมเดลล็อกลิเนียร์ที่ทดสอบมี 4 โมเดล และการทดสอบในขั้นตอนแรกมีโมเดลที่คัดออก 2 โมเดล คือ โมเดลที่ 1 และ 2 ในขั้นตอนที่สองจึงพิจารณาเฉพาะโมเดลที่ 3 และ 4 ปรากฏว่าโมเดลที่ 3 เป็นโมเดลที่กลมกลืนกับข้อมูลที่ดีที่สุด เมื่อกำหนดให้ α = .01

เพื่อความเข้าใจในการพิจารณาคัดเลือกโมเดลล็อกลิเนียร์มากขึ้น ขอยกตัวอย่างการคัดเลือกโมเดลล็อกลิเนียร์ของตัวแปร 3 ตัวแปร คือ A, B และ C โดยที่ A มี 4 ค่า B มี 2 ค่า และ C มี 2 ค่า ผลการคำนวณค่าต่าง ๆ ดังตารางที่ 9

ตารางที่ 9 ค่าสถิติอัตราส่วนไลค์ลิฮูดไค-สแควร์ และผลต่างของค่าสถิติอัตราส่วนไลค์ลิฮูดไค-สแควร์ของโมเดลล็อกลิเนียร์ กรณี 3 ตัวแปร

โมเดลที่	โมเดลแบบคูณ	H_o: Model กลมกลืนกับข้อมูล			H_o: เทอมอิทธิพลมีค่าเท่ากับศูนย์			
		L^2	DF	P	โมเดล	ΔL^2	DF	P
1	τ^a	857.97	12	.000				
2	$\tau^a\tau^b$	61.89	11	.000	1, 2	796.08	1	≤.05
3	$\tau^a\tau^b\tau^c$	19.89	10	.030	2, 3	42.00	1	≤.05
4	$\tau^a\tau^b\tau^c\tau^{ac}$	9.87	7	.196	3, 4	10.02	3	≤.05
5	$\tau^a\tau^b\tau^c\tau^{ab}$	14.15	7	.049	3, 5	5.74	3	NS
6	$\tau^a\tau^b\tau^c\tau^{bc}$	17.88	9	.037	3, 6	2.01	1	NS
7	$\tau^a\tau^b\tau^c\tau^{ac}\tau^{ab}$	4.10	4	.392	4, 7	5.77	3	NS
8	$\tau^a\tau^b\tau^c\tau^{ac}\tau^{bc}$	2.12	3	.547	4, 8	7.75	4	NS

หมายเหตุ $\alpha = .05$

จากตารางที่ 9 จะเห็นว่าในขั้นตอนที่ 1 มีโมเดลที่ 4 , 7 และ 8 ยอมรับสมมติฐานหลัก (accept H_o) ที่ระดับนัยสำคัญ .05 จึงควรนำมาทดสอบในขั้นตอนที่ 2 ซึ่งผลที่ได้ปรากฏว่าโมเดลที่ 4 เมื่อเพิ่มปฏิสัมพันธ์ของ AC เข้าไปในโมเดลที่ 3 ผลปรากฏว่าปฏิเสธสมมติฐานหลัก (reject H_o) แสดงว่าเทอมอิทธิพล AC มีนัยสำคัญทางสถิติเพียงพอที่ทำให้โมเดลที่ 4 มีความกลมกลืนกับข้อมูลเชิงประจักษ์ และเมื่อพิจารณาโมเดลที่ 7 เพิ่มปฏิสัมพันธ์ AB เข้าไปในโมเดลที่ 4 ผลปรากฏว่า ไม่สามารถปฏิเสธสมมติฐานหลัก ส่วนโมเดลที่ 8 เพิ่มปฏิสัมพันธ์ BC เข้าไปในโมเดลที่ 4 ผลปรากฏว่าไม่สามารถปฏิเสธสมมติฐานหลัก จึงสรุปได้ว่า โมเดลที่ 4 เป็นโมเดลที่เหมาะสมที่สุด

หากมีตัวแปรจำนวนมากขึ้น การสร้างตารางการณ์จรจะมีขนาดใหญ่ และมีโมเดลล็อกลิเนียร์ที่มีเทอมอิทธิพลลดหลั่นกันจำนวนมากที่นำมาพิจารณาคัดเลือก เช่น มีตัวแปร 3 ตัวแปร จะได้โมเดลจำนวน 8 โมเดล ถ้ามีตัวแปร 4 ตัวแปร จะได้โมเดลจำนวน 16 โมเดล และถ้ามีตัวแปร 5 ตัวแปร จะได้โมเดลจำนวน 64 โมเดล ฉะนั้น ยิ่งมีตัวแปรมากการทดสอบความกลมกลืนยิ่งมีความซับซ้อนมากยิ่งขึ้น เพราะการทดสอบในขั้นตอนแรกและขั้นตอนที่สอง

อาจมีโมเดลที่คาดว่าเหมาะสมจำนวนมาก จะทำให้เกิดความยากในการพิจารณาคัดเลือกโมเดล นักสถิติได้พัฒนาวิธีการที่ช่วยแก้ไข ทำให้การวิเคราะห์คัดเลือกโมเดลมีความยุ่งยากลดลง

Kennedy (1983) Ferthofer และ Lorimer (1989) สรุปว่า กระบวนการคัดเลือกโมเดล กรณีมีตัวแปรหลายตัวแปร ให้จัดกลุ่มโมเดลก่อน เช่น การวิเคราะห์ตัวแปร 3 ตัวแปร คือ A, B และ C มีโมเดลทั้งหมด 8 โมเดลที่มีลักษณะลดหลั่นกันไปเรียงตามลำดับเทอมอิทธิพลของตัวแปร A, B และ C โมเดลทั้ง 8 โมเดลจำแนกออกได้ 4 กลุ่ม แต่ละกลุ่มเรียกว่า family มีลักษณะดังนี้ กลุ่มที่ 1 มี 1 โมเดล คือ โมเดลที่ 1 เป็นกลุ่มโมเดลที่มีเทอมอิทธิพลคงที่อย่างเดียวโดยที่ไม่มีเทอมอิทธิพลใด ๆ ของตัวแปรเลย เรียกโมเดลล็อกลิเนียร์นี้ว่า โมเดล 0 ปัจจัย (0 factor) หรือ null model กลุ่มที่ 2 มี 3 โมเดล คือ โมเดลที่ 2-4 มีเทอมอิทธิพลหลัก A, B และ C เรียงกันไปตามลำดับ เช่น โมเดล 2 มีลักษณะ $F=\tau\tau^{a}$ โมเดล 3 มีลักษณะ $F=\tau\tau^{a}\tau^{b}$ และโมเดล 4 มีลักษณะ $F=\tau\tau^{a}\tau^{b}\tau^{c}$ เรียกโมเดลที่ 2-4 ว่าโมเดล 1 ปัจจัย (1 factor) หรือโมเดลปัจจัยหลัก (main marginal effect model) โมเดลที่ 4 มีเทอมอิทธิพลหลักของตัวแปรครบทุกเทอม เรียกโมเดล 4 ว่า โมเดลเต็มรูปของอิทธิพลหลัก (full main effect model) กลุ่มที่ 3 มี 3 โมเดล คือ โมเดลที่ 5-7 มีเทอมปฏิสัมพันธ์อันดับหนึ่ง (first order interaction) ได้แก่ AB, AC และ BC เพิ่มจากโมเดล 4 ทีละเทอมเรียงกันไปตามลำดับเช่นเดียวกับโมเดล 1 ปัจจัย เรียกโมเดล 5-7 ว่าโมเดล 2 ปัจจัย (2 factors) หรือโมเดลปฏิสัมพันธ์อันดับหนึ่ง (first order interaction model) โมเดลที่ 7 มีเทอมปฏิสัมพันธ์อันดับหนึ่งครบทุกเทอมจึงเรียกว่าโมเดลเต็มรูปของปฏิสัมพันธ์อันดับหนึ่ง (full first order model) และกลุ่มที่ 4 มี 1 โมเดล คือ โมเดลที่ 8 มีเทอมปฏิสัมพันธ์อันดับสอง คือ ABC (second order interaction) เพิ่มต่อจากโมเดลที่ 7 เรียกโมเดล 8 ว่าโมเดล 3 ปัจจัย (3 factors) และเนื่องจากเทอมปฏิสัมพันธ์ ABC เป็นเทอมสุดท้ายของปฏิสัมพันธ์ระหว่างตัวแปร 3 ตัว คือ A, B และ C และมีเทอมเดียว ฉะนั้นโมเดลที่ 8 จึงเป็นโมเดลเต็มรูปปฏิสัมพันธ์อันดับสอง (full second order model) และเป็นโมเดลสุดท้ายของการวิเคราะห์ตัวแปร 3 ตัวแปร จะเห็นว่าโมเดลที่ 8 มีเทอมอิทธิพลครบทั้งเทอมอิทธิพลหลักและปฏิสัมพันธ์ จึงเรียกได้อีกชื่อว่า โมเดลอิ่มตัว (saturated model) ส่วนโมเดลที่ 1 - 7 มีเทอมอิทธิพลไม่ครบทุกเทอม เรียก 7 โมเดลนี้ว่าโมเดลไม่อิ่มตัว (unsaturated model)

การคัดเลือก 8 โมเดลนี้มีวิธีการลดจำนวนโมเดลในการทดสอบได้โดยการเลือกโมเดลเต็มรูปของแต่ละกลุ่ม คือ โมเดลที่ 1, 4, 7 และ 8 มาทำการทดสอบสมมติฐานของความกลมกลืนของโมเดลตาม 2 ขั้นตอนที่กล่าวแล้ว เมื่อโมเดลใดเป็นโมเดลที่เหมาะสม

จึงคัดเลือกโมเดลในกลุ่มเดียวกันอีกครั้งด้วยการทดสอบความกลมกลืนตามกระบวนการคัดเลือก 2 ขั้นตอน วิธีการนี้จะทำให้ง่ายต่อการพิจารณามากขึ้น วิธีการคัดเลือกนี้สามารถนำไปพัฒนากับการวิเคราะห์ตัวแปรมากกว่า 3 ได้ง่ายและสะดวกขึ้น

อย่างไรก็ตาม กระบวนการคัดเลือกโมเดล 2 ขั้นตอนนี้สามารถใช้กับการศึกษาทั้งแบบสมมาตรและแบบอสมมาตร แต่มีข้อแตกต่างกัน 3 ประการ ประการแรก โมเดลในการพิจารณาสำหรับการศึกษาแบบสมมาตรเป็นโมเดลล็อกลิเนียร์ทุกโมเดลที่เป็นไปได้ที่ใช้ทำนายความถี่ที่คาดหวัง เรียกว่า โมเดลล็อกลิเนียร์ทั่ว ๆ ไป (general log-linear model) แต่โมเดลในการพิจารณาสำหรับการศึกษาแบบอสมมาตร นอกจากมีโมเดลล็อกลิเนียร์ทั่ว ๆ ไปแล้ว ยังมีโมเดลที่พัฒนาจากโมเดลล็อกลิเนียร์ทั่ว ๆ ไป เรียกว่า โมเดลโลจิท (logit model) โมเดลโลจิทแสดงความสัมพันธ์ระหว่างผลต่างของสัดส่วนของความถี่ในระดับของตัวแปรตามสองระดับกับชุดอิทธิพลของตัวแปรต้น ประการที่สอง การศึกษาแบบอสมมาตรมีโมเดลสำหรับพิจารณาน้อยกว่าการศึกษาแบบสมมาตร และประการที่สาม การเลือกโมเดลในการศึกษาแบบอสมมาตรไม่เพียงแต่มีโมเดลที่เป็นโมเดลประหยัดและมีความกลมกลืนกับข้อมูลเชิงประจักษ์แต่โมเดลที่เลือกจำเป็นต้องเริ่มพิจารณาจากโมเดลโลจิทแบบศูนย์ (null logit model) เช่น โมเดล 4 โมเดลของการศึกษา 2 ตัวแปรดังตารางที่ 8 หน้า 27 โมเดลที่ 1-4 เป็นโมเดลสำหรับการศึกษาแบบสมมาตร และโมเดลที่ 3 และ 4 เป็นโมเดลสำหรับการศึกษาแบบอสมมาตร เมื่อกำหนดให้ตัวแปร A เป็นตัวแปรต้น และ B เป็นตัวแปรตาม เรียกตัวแปร B ว่าตัวแปรโลจิท (logit variable) และโมเดล 3 เป็นโมเดลที่ไม่มีเทอมปฏิสัมพันธ์ของ A และ B แสดงว่า ตัวแปร A ไม่มีอิทธิพลต่อตัวแปร B จึงเรียกว่าโมเดลโลจิทแบบศูนย์ (null logit model) ส่วนโมเดล 4 เป็นโมเดลที่แสดงว่า A มีความสัมพันธ์กับ B และ A มีโอกาสที่จะมีอิทธิพลต่อ B เมื่อทดสอบนัยสำคัญของเทอมปฏิสัมพันธ์ AB แล้วปรากฏว่ามีนัยสำคัญทางสถิติเพียงพอที่ทำให้โมเดลมีความกลมกลืนกับข้อมูล แสดงว่าตัวแปร A มีผลต่อตัวแปร B

จากโมเดลจำนวนมากเมื่อผ่านกระบวนการคัดเลือกโมเดล กรณีการศึกษาแบบสมมาตร การคัดเลือกโมเดลควรพิจารณาคัดเลือกให้เหลือเพียงโมเดลเดียวที่กลมกลืนกับข้อมูลมากที่สุด แต่การศึกษาแบบอสมมาตรไม่จำเป็นต้องมีโมเดลเดียวที่มีความกลมกลืนกับข้อมูลเมื่อได้โมเดลที่คัดเลือกไว้แล้ว ขั้นตอนต่อไปของการวิเคราะห์ล็อกลิเนียร์ คือ การตรวจสอบโมเดลว่ามีความเหมาะสมหรือไม่ และการประมาณค่าพารามิเตอร์ลอการิทึมของอัตราส่วนแต้มต่อเพื่อการแปลผลต่อไป สำหรับการตรวจสอบโมเดลมีรายละเอียดดังหัวข้อต่อไป

5. การตรวจสอบความเหมาะสมของโมเดล (Examine of appropriated model)

เมื่อคัดเลือกโมเดลที่มีความกลมกลืนกับข้อมูลเชิงประจักษ์แล้วงานขั้นต่อไป คือ การตรวจสอบ เพื่อความมั่นใจว่าเป็นโมเดลที่เป็นตัวแทนที่ดีที่ใช้ทำนายความถี่ที่คาดหวังและนำไปสู่การสรุปผลการศึกษาได้ถูกต้อง การวิเคราะห์ล็อกลิเนียร์มีวิธีการตรวจสอบ 2 แบบ คือ แบบที่ 1 ตรวจสอบด้วยการพิจารณาค่าเศษเหลือมาตรฐาน (standardize residual , r_{ij}) และแบบที่ 2 ตรวจสอบด้วยการทดสอบนัยสำคัญของเทอมอิทธิพล

แบบที่ 1 การพิจารณาค่าเศษเหลือมาตรฐาน (standardize residual , r_{ij})

การคำนวณค่าเศษเหลือมาตรฐาน (r_{ij}) จากสูตร

$$r_{ij} = \frac{(f_{ij} - F_{ij})}{\sqrt{F_{ij}}}$$

เกณฑ์สำหรับพิจารณาคือ ถ้า ค่า $r_{ij} < 2$ แสดงว่าโมเดลล็อกลิเนียร์ที่คัดเลือกมีความกลมกลืนกับข้อมูลเชิงประจักษ์ได้อย่างเหมาะสม

แบบที่ 2 การทดสอบนัยสำคัญของเทอมอิทธิพล

เป็นการทดสอบนัยสำคัญของเทอมอิทธิพลแต่ละเทอมในโมเดลที่คัดเลือกว่ามีเพียงพอที่ทำให้โมเดลมีความกลมกลืนกับข้อมูล สถิติทดสอบ คือ z- test โดยมีสมมติฐานหลัก คือ เทอมอิทธิพลมีค่าเท่ากับศูนย์ เทอมอิทธิพลในที่นี้อยู่ในรูปของ λ เนื่องจากโปรแกรมคอมพิวเตอร์ปัจจุบันมีพื้นฐานการคำนวณจากโมเดลแบบบวก เช่น การวิเคราะห์ความแปรปรวนและการวิเคราะห์การถดถอย ฉะนั้น โมเดลล็อกลิเนียร์ปกติอยู่ในรูปโมเดลแบบคูณสามารถปรับให้อยู่ในรูปโมเดลแบบบวกได้โดยใช้ลอการิทึม การทดสอบนัยสำคัญของอิทธิพลจึงนิยมใช้เทอมของ λ และเมื่อกำหนดให้ $H_0 : \lambda = 0$ และสูตรในการคำนวณ z คือ $z = \frac{\lambda}{SE_\lambda}$

ยกตัวอย่างการทดสอบนัยสำคัญของอิทธิพลหลักของตัวแปร A และปฏิสัมพันธ์อันดับหนึ่ง AB สำหรับการวิเคราะห์ตัวแปร 2 ตัวแปร คือ A และ B โดยที่ตัวแปร A มี a ค่า และ B มี b ค่า การทดสอบนัยสำคัญของเทอมอิทธิพลหลักของ A และปฏิสัมพันธ์อันดับหนึ่ง AB โดยมีสมมติฐานหลัก ดังนี้

การทดสอบนัยสำคัญของเทอมอิทธิพลหลัก A

H_0: $\lambda^a_i = 0$, $i = 1, 2, ..., a$

สถิติที่ทดสอบได้แก่ z-test และ $z = \frac{\lambda^a_i}{SE_{\lambda^a_i}}$

โดย $$SE\lambda_i^{*} = \frac{\left(\sum_{i=1}^{a}\sum_{j=1}^{b}\frac{1}{f_{ij}}\right)^{\frac{1}{2}}}{a}$$

สำหรับการทดสอบนัยสำคัญของเทอมอิทธิพลปฏิสัมพันธ์ AB

H_o: $\lambda_{ij}^{ab} = 0$, $i = 1, 2, \ldots, a$ $j = 1, 2, \ldots, b$

สถิติที่ทดสอบ คือ $$z = \frac{\lambda_{ij}^{ab}}{SE\lambda_{ij}^{ab}}$$

โดย $$SE\lambda_{ij}^{ab} = \frac{\left(\sum_{i=1}^{a}\sum_{j=1}^{b}\frac{1}{f_{ij}}\right)^{\frac{1}{2}}}{ab}$$

เกณฑ์ในการพิจารณา คือ ค่า z ที่คำนวณได้ต้องมากกว่า 1.96 เมื่อกำหนดระดับนัยสำคัญ 5 เปอร์เซ็นต์ หรือ เมื่อพิจารณาจากช่วงเชื่อมั่นในการประมาณค่าพารามิเตอร์ที่ระดับความเชื่อมั่น 95 เปอร์เซ็นต์ ค่าประมาณพารามิเตอร์(λ) ต่ำสุดและสูงสุด หากไม่ครอบคลุมค่าศูนย์จะทำให้ผลการทดสอบปฏิเสธสมมติฐานหลัก แสดงว่า ค่าประมาณพารามิเตอร์ (λ) มีค่าไม่เท่ากับศูนย์ หมายความว่า เทอมอิทธิพลมีนัยสำคัญเพียงพอต่อการอธิบายหรือการทำนายค่าความถี่ที่คาดหวัง ฉะนั้น เทอมอิทธิพลนี้จึงควรมีอยู่ในโมเดลและยังคงทำให้โมเดลมีความกลมกลืนกับข้อมูลเชิงประจักษ์ได้อย่างเหมาะสม

จากที่กล่าวทั้งหมด หากผู้วิจัยศึกษาความสัมพันธ์แบบสมมาตรที่เป็นการศึกษาความสัมพันธ์ระหว่างตัวแปรโดยมิได้กำหนดตัวแปรใดเป็นตัวแปรตามและตัวแปรต้น เมื่อผลการคัดเลือกโมเดลและตรวจสอบโมเดลสรุปได้ว่าตัวแปรมีความสัมพันธ์กัน จะต้องทำการศึกษาต่อไปว่า ความสัมพันธ์นั้นมีขนาด และทิศทางอย่างไรด้วยวิธีการวัดความสัมพันธ์ (measure of association) เช่น สถิติ Q ของยูล (Yule's Q), สัมประสิทธิ์การณ์จร (contingency coefficient), สถิติวีกำลังสองของเครเมอร์ (Cramer's V^2) อัตราส่วนแต้มต่อ (odds ratio) หรืออื่น ๆ สถิติดังกล่าวเลือกใช้ตามข้อจำกัดของแต่ละวิธี แต่ถ้าสรุปได้ว่าตัวแปรไม่มีความสัมพันธ์กันไม่จำเป็นต้องมีการศึกษาหาขนาด และทิศทางต่อไป หากผู้วิจัยศึกษาความสัมพันธ์แบบอสมมาตรที่กำหนดให้ตัวแปรหนึ่งเป็นตัวแปรตามและตัวแปรอื่น ๆ เป็นตัวแปรต้น และสรุปได้ว่า ตัวแปรต้นใดมีอิทธิพลต่อตัวแปรตาม ขั้นต่อไปเป็นการประมาณค่าพารามิเตอร์

ลอการิทึมของอัตราส่วนแต้มต่อ คือ ค่าผลต่างของโอกาสในการเกิดเหตุการณ์สองเหตุการณ์ที่ใช้อธิบายความแตกต่างของความถี่ 2 ระดับของตัวแปรตาม และสะท้อนให้ทราบถึงขนาดอิทธิพลที่ส่งผลต่อตัวแปรตาม การประมาณค่าลอการิทึมอัตราส่วนแต้มต่อ มีรายละเอียดดังหัวข้อต่อไป

6. การประมาณค่าพารามิเตอร์ลอการิทึมของอัตราส่วนแต้มต่อ (Log Odds Ratio)

ค่าลอการิทึมของอัตราส่วนแต้มต่อ คือ ค่าผลต่างของโอกาสในการเกิดเหตุการณ์สองเหตุการณ์ที่ใช้อธิบายความแตกต่างของความถี่ 2 ระดับของตัวแปรตาม ในที่นี้เป็นค่าที่บ่งบอกถึงค่าของอิทธิพลของตัวแปรหนึ่งที่มีต่ออีกตัวแปรหนึ่ง การประมาณค่าลอการิทึมของอัตราส่วนแต้มต่อประมาณค่าได้จากโมเดลโลจิทที่พัฒนาจากโมเดลล็อกลิเนียร์ทั่ว ๆ ไปที่ได้คัดเลือกว่าเป็นโมเดลที่มีความกลมกลืนกับข้อมูลอย่างเหมาะสม ยกตัวอย่างเช่น การศึกษา 2 ตัวแปร คือ A และ B โดยที่ A มี 2 ค่า และ B มี 2 ค่า โมเดลล็อกลิเนียร์จะมี 4 โมเดล ตามตารางที่ 8 หน้า 27 คือ โมเดลที่ 1-4 และโมเดลสำหรับการศึกษาความสัมพันธ์แบบอสมมาตรมี 2 โมเดล คือ โมเดลที่ 3 และ 4 เมื่อกำหนดให้ B เป็นตัวแปรตาม และ A เป็นตัวแปรต้น หากโมเดล 4 มีความกลมกลืนกับข้อมูล แสดงว่า A มีอิทธิพลต่อ B และโมเดล 4 มีลักษณะโมเดลแบบคูณ ดังนี้

$$F_{ij} = \tau\tau_i^{a}\tau_j^{b}\tau_{ij}^{ab} \qquad \text{.....(1)}$$

สามารถนำโมเดลนี้มาพัฒนาเป็นโมเดลโลจิทเพื่อประมาณค่าอิทธิพลของ A ในที่นี้ คือ ค่าลอการิทึมของอัตราส่วนแต้มต่อของตัวแปร A ตามที่ Goodman (1972) เสนอไว้และอธิบายโดย Knoke และ Burke (1980) สรุปได้ว่า โมเดลโลจิทมีเกณฑ์ที่ใช้ในการพิจารณา คือ สัดส่วนของความถี่ที่คาดหวัง 2 ระดับของตัวแปรตาม ซึ่งต่างจากโมเดลล็อกลิเนียร์ทั่ว ๆ ไป ที่ใช้ความถี่ที่คาดหวังเป็นเกณฑ์ในการพิจารณา ยกตัวอย่าง เช่น กรณีโมเดลตามสมการที่ (1) พัฒนาเป็นโมเดลโลจิท ดังนี้

$$\frac{F_{i1}}{F_{i2}} = \frac{\tau\tau_i^{a}\tau_1^{b}\tau_{i1}^{ab}}{\tau\tau_i^{a}\tau_2^{b}\tau_{i2}^{ab}}$$

$$= \frac{\tau_1^{b}\tau_{i1}^{ab}}{\tau_2^{b}\tau_{i2}^{ab}}$$

เมื่อพิจารณาตามข้อตกลงจะได้ $\tau_1^{b}\tau_2^{b} = 1$ และ $\tau_{i1}^{ab}\tau_{i2}^{ab} = 1$

ฉะนั้น $F_{i1}/F_{i2} = \tau_1^{b}\tau_1^{b}\tau_{i1}^{ab}\tau_{i1}^{ab}$

$$\ln(F_{i1}/F_{i2}) = \ln\tau_1^{b} + \ln\tau_1^{b} + \ln\tau_{i1}^{ab} + \ln\tau_{i1}^{ab}$$

$$= 2\ln\tau_1^{b} + 2\ln\tau_{i1}^{ab}$$

$$= 2\lambda_1^{b} + 2\lambda_{11}^{ab}$$

$$\Phi_1^{b} = \beta_1^{b} + \beta_{11}^{ab} \quad \text{.....(2)}$$

เมื่อ $\Phi_1^{b} = \ln(F_{11} / F_{12})$

$\beta_1^{b} = 2\lambda_1^{b}$

$\beta_{11}^{ab} = 2\lambda_{11}^{ab}$

สมการ (2) เรียกว่าโมเดลโลจิท โดย $-\infty \leq \beta_1^{b} \leq +\infty$, $-\infty \leq \beta_{11}^{ab} \leq +\infty$ และ $\beta_1^{b} + \beta_2^{b} = 0$, $\beta_{11}^{ab} + \beta_{21}^{ab} = 0$ จะเห็นว่า ค่า β_1^{b} เป็นค่าที่ประมาณได้จากโมเดลโลจิท เรียกว่า ลอการิทึมของอัตราส่วนแต้มต่อ หากค่าที่ได้มีค่าบวก แสดงว่าอิทธิพลของเทอมนั้น ๆ มีผลทำให้การทำนายค่า Φ_1^{b} มีค่าเพิ่มมากขึ้น และหากมีค่าลบ แสดงว่า อิทธิพลของเทอมนั้น ๆ มีผลทำให้การทำนายค่า Φ_1^{b} มีค่าน้อยลง (Knoke และ Burke, 1980; Norusis, 1992)

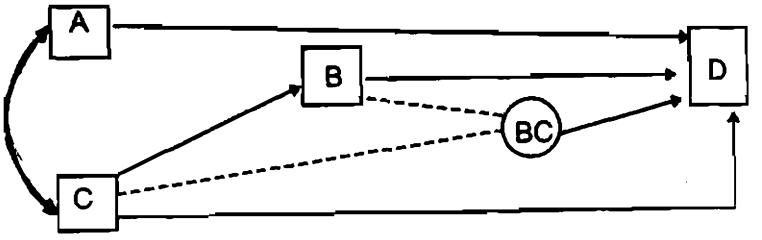
อย่างไรก็ตาม เมื่อประมาณค่าลอการิทึมอัตราส่วนแต้มต่อแล้วต้องเปลี่ยนค่าเป็นอัตราส่วนแต้มต่อเพื่อการแปลผลต่อไป เนื่องจากค่าอัตราส่วนแต้มต่อ คือ สัดส่วนของโอกาสในการเกิดเหตุการณ์สองเหตุการณ์ ฉะนั้น อัตราส่วนแต้มต่อมีค่ามากกว่าศูนย์เสมอ ถ้าอัตราส่วนแต้มต่อมีค่าเท่ากับ 1 แสดงว่า โอกาสของการเกิดเหตุการณ์ 2 เหตุการณ์มีเท่า ๆ กัน ถ้าอัตราส่วนแต้มต่อมีค่ามากกว่า 1 แสดงว่า โอกาสการเกิดเหตุการณ์หนึ่งมีมากกว่าอีกเหตุการณ์หนึ่ง และถ้าอัตราส่วนแต้มต่อมีค่าน้อยกว่า 1 แสดงว่า โอกาสการเกิดเหตุการณ์หนึ่งมีน้อยกว่าอีกเหตุการณ์หนึ่ง (Knoke และ Burke, 1980; Norusis, 1992)

ตอนที่ 2 การประยุกต์โมเดลล็อกลิเนียร์ในการวิเคราะห์สาเหตุ (Causal Analysis)

การวิเคราะห์สาเหตุเป็นการศึกษาความสัมพันธ์ระหว่างตัวแปรที่สนใจ เพื่อสรุปความสัมพันธ์เชิงสาเหตุระหว่างตัวแปร การวิจัยลักษณะนี้แบ่งออกเป็น 2 ประเภท คือ 1) การวิเคราะห์สาเหตุในสภาพการทดลอง เป็นการศึกษาความสัมพันธ์เชิงสาเหตุระหว่างตัวแปรภายใต้การออกแบบสภาพการทดลองที่มีการควบคุมตัวแปรแทรกซ้อน มีการสุ่มกลุ่มตัวอย่างสุ่มการจัดกลุ่มทดลองและกลุ่มควบคุม ตลอดจนสุ่มสิ่งจัดกระทำ (treatment) และ 2) การวิเคราะห์สาเหตุในสภาพที่ไม่ใช่การทดลอง เป็นการศึกษาความสัมพันธ์เชิงสาเหตุระหว่างตัวแปรภายใต้สถานการณ์ธรรมชาติ เช่น การศึกษาอิทธิพลหรือสาเหตุของการเกิดเหตุการณ์หรือ

การเปลี่ยนแปลงค่าของตัวแปรที่น่าสนใจโดยมีการควบคุมตัวแปรโดยวิธีการทางสถิติ (ศิริชัย กาญจนวาสี, ทวีวัฒน์ ปิตยานนท์ และดิเรก ศรีสุโข, 2537)

สำหรับการวิจัยนี้เป็นการวิเคราะห์สาเหตุในสภาพที่ไม่ใช่การทดลองโดยมุ่งศึกษาความสัมพันธ์เชิงสาเหตุแบบทางเดียว (recursive causal relations) ระหว่างตัวแปรที่ส่งผลต่อความสำเร็จในการทำวิทยานิพนธ์ของนิสิตมหาบัณฑิตทางสังคมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ปีการศึกษา 2535 และ 2536 โดยประยุกต์ใช้กับโมเดลล็อกลิเนียร์ การประยุกต์ทำได้โดยพิจารณาในเรื่องของการวิเคราะห์อิทธิพล (path analysis) เพื่อศึกษาอิทธิพลทางตรง (direct effects) และอิทธิพลทางอ้อม (indirect effects) ของตัวแปรที่มีผลต่อตัวแปรที่สนใจ แนวความคิดพื้นฐานในการวิเคราะห์อิทธิพล คือ ตัวแปรมีความสัมพันธ์เชิงสาเหตุด้วยหลัก 3 ประการ คือ 1) การแยก (isolation) 2) ความเกี่ยวเนื่อง (association) และ 3) ลำดับการเกิดก่อน (temporal order) ตัวอย่างเช่น ถ้าเราสามารถแยกให้เห็นได้ว่าตัวแปร X เป็นตัวแปรเพียงตัวแปรเดียวที่เกิดก่อน มีความเกี่ยวเนื่องและมีผลทำให้เกิดความแตกต่างในตัวแปร Y จะสรุปได้ว่าตัวแปร X เป็นสาเหตุของตัวแปร Y โดยทั่วไปตัวแปรที่ใช้ในการศึกษาความสัมพันธ์ แบ่งเป็น 2 ประเภท คือ ตัวแปรภายนอก (exogenous variable) และตัวแปรภายใน (endogenous variable) ตัวแปรภายนอก เป็นตัวแปรที่นักวิจัยไม่สนใจศึกษาสาเหตุของตัวแปรเหล่านี้ ส่วนตัวแปรภายใน เป็นตัวแปรที่นักวิจัยสนใจศึกษาว่าได้รับอิทธิพลจากตัวแปรใด สาเหตุของตัวแปรภายในจะแสดงไว้ในโมเดลอย่างชัดเจน (นงลักษณ์ วิรัชชัย, 2538; Goodman, 1973; Knoke และ Burke, 1980; Kennedy, 1983) จากแนวคิดนี้ Goodman (1973) ได้ประยุกต์ใช้โมเดลล็อกลิเนียร์ในการวิเคราะห์อิทธิพลของตัวแปร 4 ตัวแปรซึ่งหลักในการวิเคราะห์คล้ายกับที่ Kennedy (1983) ได้กล่าวถึงการประยุกต์ใช้โมเดลล็อกลิเนียร์ในการวิเคราะห์ความสัมพันธ์เชิงสาเหตุระหว่างตัวแปร 4 ตัวแปรดังตัวอย่างโครงสร้างความสัมพันธ์ระหว่างตัวแปร A, B, C และ D ตามสมมติฐานที่ใช้เป็นแนวทางในการพิจารณาประยุกต์ใช้โมเดลล็อกลิเนียร์ตามรูปภาพที่ 1 ในที่นี้ตัวแปรหลักแทนด้วยสัญญลักษณ์กรอบรูปสี่เหลี่ยม ปฏิสัมพันธ์ระหว่างตัวแปรแทนด้วยวงกลม สัญญลักษณ์ลูกศรแทนความสัมพันธ์เชิงสาเหตุระหว่างตัวแปรโดยหัวลูกศรชี้ไปที่ตัวแปรผล ส่วนเส้นโค้งและมีลูกศร 2 หัวแทนความสัมพันธ์ระหว่างตัวแปร หมายถึง ตัวแปรแต่ละตัวแปรต่างก็เป็นเหตุและผลซึ่งกันและกัน และเส้นประ แทน เส้นเชื่อมโยงระหว่างเทอมที่มีปฏิสัมพันธ์กัน โดยไม่ได้หมายถึงตัวแปรส่งผลต่อเทอมปฏิสัมพันธ์



รูปภาพที่ 1 โมเดลความสัมพันธ์เชิงสาเหตุแบบทางเดียวของตัวแปร A, B, C, D ตามสมมติฐาน

จากรูปภาพที่ 1 Kennedy (1983) อธิบายโครงสร้างความสัมพันธ์ของตัวแปร 4 ตัวแปร คือ A , B , C และ D สรุปได้ว่าตัวแปร 4 ตัวแปรจะมี 2 ตัวแปรที่เป็นตัวแปรภายใน (endogenous variables) คือ B และ D และเมื่อพิจารณาลำดับการเกิดของตัวแปรจากซ้ายไปขวาจะเห็นว่า A และ C มีความสัมพันธ์กัน และเป็นตัวแปรที่เกิดก่อน B โดย B ได้รับอิทธิพลทางตรงจากตัวแปร C และ B เป็นตัวแปรที่เกิดก่อน D และ D ได้รับอิทธิพลทางตรงจาก A , B และปฏิสัมพันธ์อันดับสอง BC และรับอิทธิพลทั้งทางตรงและทางอ้อมจาก C จากลักษณะโครงสร้างความสัมพันธ์นี้เป็นโครงสร้างที่ได้มาจากทฤษฎี ผู้วิจัยต้องดำเนินการตรวจสอบว่าเป็นไปตามทฤษฎีหรือไม่ ด้วยความพยายามของนักสถิติที่จะประยุกต์การวิเคราะห์ล็อกลิเนียร์กับการวิเคราะห์อิทธิพล นักสถิติจึงคิดหาวิธีเพื่อการทดสอบตามทฤษฎีซึ่งเทียบเคียงได้กับการวิเคราะห์การถดถอยในการวิเคราะห์อิทธิพลโดยสร้างโมเดลล็อกลิเนียร์สำหรับตัวแปรผลจากซ้ายไปขวาและตัวแปรเหตุของตัวแปรผลนั้น ๆ โดยพิจารณาถึงความกลมกลืนของโมเดลกับข้อมูลเชิงประจักษ์และทดสอบนัยสำคัญของเทอมอิทธิพลแต่ละเทอมในโมเดลล็อกลิเนียร์นั้น ๆ หลังจากขจัดอิทธิพลอื่น ๆ ออก อย่างไรก็ตาม การประยุกต์การวิเคราะห์ล็อกลิเนียร์สามารถวิเคราะห์ความสัมพันธ์เชิงสาเหตุของตัวแปรที่มีลักษณะความสัมพันธ์ในทิศทางเดียวเท่านั้น

การพิจารณาสร้างโมเดลล็อกลิเนียร์ตามเส้นทางความสัมพันธ์ของตัวแปรตามรูปภาพที่ 1 จะสามารถพิจารณาสร้างโมเดลล็อกลิเนียร์เพื่อการวิเคราะห์ได้ 3 ขั้นตอน คือ ขั้นตอนแรกสร้างตาราง 2 มิติ เพื่อวิเคราะห์ความสัมพันธ์ระหว่าง 2 ตัวแปร คือ A และ C ก่อน ขั้นตอนที่สองสร้างตาราง 3 มิติเพื่อวิเคราะห์ความสัมพันธ์เชิงสาเหตุระหว่างตัวแปร A, B และ C โดยกำหนดให้ตัวแปร B เป็นตัวแปรผลหรือตัวแปรตาม ขั้นตอนที่สาม สร้างตาราง 4 มิติเพื่อวิเคราะห์ความสัมพันธ์เชิงสาเหตุระหว่างตัวแปร A , B , C และ D โดยกำหนดให้ตัวแปร D เป็นตัวแปรผลหรือตัวแปรตาม และแต่ละขั้นตอนดำเนินการวิเคราะห์ตามกระบวนการวิเคราะห์ล็อกลิเนียร์

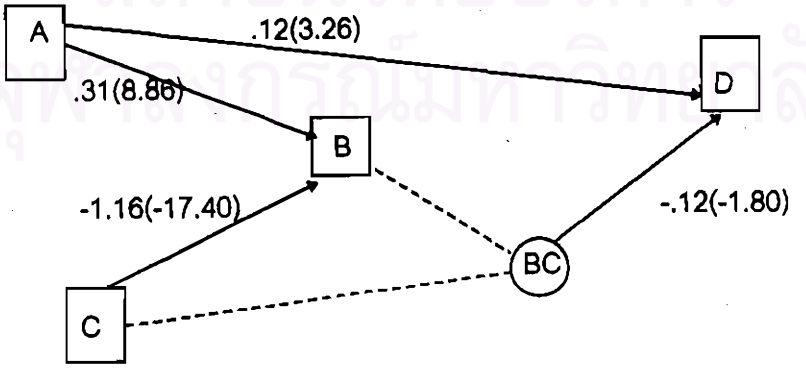
3 กระบวนการ คือ 1) สร้างโมเดลล็อกลิเนียร์ 2) ทดสอบความกลมกลืนของโมเดลล็อกลิเนียร์ และ 3) ตรวจสอบโมเดล และประมาณอิทธิพล จากกระบวนการต่าง ๆ Kennedy (1983) ได้แสดงผลการวิเคราะห์ดังตารางที่ 10 และโมเดลความสัมพันธ์ของตัวแปร A, B, C และ D พร้อมค่าอิทธิพลและค่าสถิติ z ดังรูปภาพที่ 2 สรุปได้ว่า ตัวแปร D ได้รับเฉพาะอิทธิพลทางตรงของตัวแปรหลัก A และปฏิสัมพันธ์อันดับหนึ่งของ B กับ C เท่านั้น และ Kennedy (1983) แสดงค่าสัมประสิทธิ์เส้นทางอิทธิพลด้วยค่าอิทธิพลในเทอมของ λ ไม่ได้เสนอการแปลผล อย่างไรก็ตาม Knoke และ Burke (1980) ได้เสนอค่าสัมประสิทธิ์เส้นทางอิทธิพลด้วยค่าลอการิทึมของอัตราส่วนแต้มต่อที่ประมาณได้จากโมเดลโลจิท ซึ่งสามารถอธิบายอิทธิพลของตัวแปรได้เช่นเดียวกับค่าสัมประสิทธิ์การถดถอยในรูปคะแนนมาตรฐานของการวิเคราะห์การถดถอย และสำหรับการวิเคราะห์ล็อกลิเนียร์สามารถแปลผลการวิเคราะห์ได้ชัดเจนมากขึ้นในรูปของอัตราส่วนแต้มต่อ ซึ่งเปลี่ยนค่ามาจากค่าลอการิทึมของอัตราส่วนแต้มต่อ

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 10 ค่าผลต่างของอัตราส่วนไลค์ลิฮูดไค-สแควร์ของโมเดลล็อกลิเนียร์ตามโมเดลสมมติฐานรูปภาพที่ 1

ขนาดตาราง	โมเดลที่	ลักษณะโมเดลตามฐานข้อมูล	โมเดลคู่ทดสอบ	ΔL^2	df	p
2 มิติ	1	[A][C]				
	2	[AC]	1, 2	86.04	1	.00
3 มิติ	3	[AC][BC]				
	4	[AC][BC][AB]	3, 4	30.83	1	.00
	5	[AC][AB]				
	6	[AC][AB][BC]	5, 6	591.05	1	.00
	7	[ABC]	6, 7	1.27	1	.26
4 มิติ	8	[ABC][BD][CD]				
	9	[ABC][BD][CD][AD]	8, 9	9.37	1	.00
	10	[ABC][AD][CD]				
	11	[ABC][AD][CD][BD]	10, 11	0.12	1	.72
	12	[ABC][AD][BD]				
	13	[ABC][AD][BD][CD]	12, 13	2.40	1	.12
	14	[ABC][ABD][ACD]				
	15	[ABC][ABD][ACD][BCD]	14, 15	2.90	1	.04
	16	[ABCD]				

หมายเหตุ $\alpha = .01$



รูปภาพที่ 2 โมเดลความสัมพันธ์ของตัวแปร A , B , C , D

ตอนที่ 3 งานวิจัยที่เกี่ยวข้อง

ผู้วิจัยเสนอสาระในตอนนี้เป็น 4 หัวข้อ คือ 1. ความหมายและความสำคัญของวิทยานิพนธ์ 2. ผลงานวิจัยเกี่ยวกับการทำวิทยานิพนธ์และตัวแปรที่มีผลต่อระยะเวลาที่ใช้ในการศึกษาต่างกัน 3. งานวิจัยเกี่ยวกับการประยุกต์โมเดลล็อกลิเนียร์ และ 4. สมมติฐานวิจัย ดังรายละเอียดต่อไปนี้

1. ความหมายและความสำคัญของวิทยานิพนธ์

การศึกษาในระดับบัณฑิตศึกษาของทุกมหาวิทยาลัย นิสิต นักศึกษาแผน ก หรือนิสิต นักศึกษาที่เรียนแผนการเรียนเน้นการทำวิจัยทุกคนต้องเสนอรายงานผลงานวิชาการที่เรียกว่า วิทยานิพนธ์ซึ่งเป็นส่วนหนึ่งของหลักสูตร (ปรีชา คัมภีรปกรณ์, สิริวรรณ ศรีพหล และปรีชา วิหคโต, 2537) การทำวิทยานิพนธ์จึงถือได้ว่าเป็นบทเรียนเริ่มต้นในการทำวิจัยของนิสิต

คำว่า วิทยานิพนธ์ ตามความหมายในพจนานุกรม ฉบับราชบัณฑิตยสถาน (2525) ระบุไว้ว่าวิทยานิพนธ์ หมายถึง บทนิพนธ์ที่ผู้เรียบเรียงยกเอาหัวเรื่องใดเรื่องหนึ่งขึ้นวิจัยและพัฒนาขยายความเพื่อเสนอรับปริญญา ปรีชา คัมภีรปกรณ์, สิริวรรณ ศรีพหล และปรีชา วิหคโต (2537) ได้ให้ความหมายว่า วิทยานิพนธ์ หมายถึง เอกสารรายงานการค้นคว้าวิจัย หรือทดลองเพื่อขอรับปริญญาในแขนงวิชาใดวิชาหนึ่ง หรือเป็นรายงานทางวิชาการในการค้นคว้าวิจัยอย่างเป็นระบบ ผลจากการทำวิทยานิพนธ์ทำให้เกิดความรู้ใหม่ หรือแนวคิดใหม่ในแขนงวิชานั้น ๆ จุดมุ่งหมายสำคัญของการทำวิทยานิพนธ์ คือ การฝึกฝนให้นิสิต นักศึกษารู้จักวิธีการค้นคว้าอย่างมีเหตุผลและมีระเบียบ ส่งเสริมให้สามารถใช้ภาษาได้ถูกต้องและชัดเจน

ความสำคัญของวิทยานิพนธ์ในการศึกษาระดับบัณฑิตศึกษาของประเทศไทยสะท้อนให้เห็นได้จากจำนวนหน่วยกิตที่กำหนดขึ้นตามเกณฑ์มาตรฐานของทบวงมหาวิทยาลัย คือ

หลักสูตรระดับปริญญาโท ให้มีจำนวนหน่วยกิตรวมตลอดหลักสูตรไม่น้อยกว่า 36 หน่วยกิต โดยแบ่งการศึกษาเป็น 2 แผน คือ

แผน ก เป็นแผนการศึกษาที่เน้นการวิจัย โดยมีการทำวิทยานิพนธ์ การศึกษาตามแผน ก มี 2 แบบ คือ

แบบ ก(1) ทำเฉพาะวิทยานิพนธ์ ซึ่งมีค่าเทียบได้ไม่น้อยกว่า 36 หน่วยกิต โดยสถาบันอุดมศึกษาอาจกำหนดให้เรียนรายวิชาเพิ่มเติมหรือทำ

กิจกรรมทางวิชาการอื่นเพิ่มขึ้นก็ได้ โดยไม่นับหน่วยกิต แต่จะต้องมีผลสัมฤทธิ์ตามที่สถาบันกำหนด

แบบ ก(2) ทำวิทยานิพนธ์ซึ่งมีค่าเทียบได้ไม่น้อยกว่า 12 หน่วยกิต และต้องศึกษางานรายวิชาอีกไม่น้อยกว่า 12 หน่วยกิต

แผน ข เป็นแผนการศึกษาที่เน้นการศึกษารายวิชา โดยไม่ต้องทำวิทยานิพนธ์ แต่ต้องมีการศึกษาค้นคว้าด้วยตนเองไม่น้อยกว่า 3 หน่วยกิต แต่ไม่เกิน 6 หน่วยกิต '

จะเห็นว่าหน่วยกิตของวิทยานิพนธ์มีตั้งแต่ 12-36 หน่วยกิต เมื่อจำนวนหน่วยกิตตลอดหลักสูตรไม่น้อยกว่า 36 หน่วยกิตในแบบ ก แสดงให้เห็นว่าวิทยานิพนธ์มีความสำคัญต่อการเรียนในระดับปริญญาโทมาก

2. ผลงานวิจัยเกี่ยวกับการทำวิทยานิพนธ์และตัวแปรที่มีผลต่อระยะเวลาที่ใช้ในการศึกษาต่างกัน

งานวิจัยที่ศึกษาเกี่ยวกับการทำวิทยานิพนธ์ แบ่งได้เป็น 3 กลุ่ม กลุ่มแรกเป็นงานวิจัยเพื่อวิเคราะห์เนื้อหาวิทยานิพนธ์ กลุ่มที่สองเป็นงานวิจัยที่ศึกษาตัวแปรที่ส่งผลต่อระยะเวลาที่ใช้ในการศึกษา และกลุ่มที่สามเป็นงานวิจัยที่ศึกษาตัวแปรที่ส่งผลต่อความสำเร็จในการทำวิทยานิพนธ์ ดังนี้

2.1 งานวิจัยเพื่อวิเคราะห์เนื้อหาวิทยานิพนธ์ของนิสิตโดยเฉพาะวิทยานิพนธ์ของนิสิตจุฬาลงกรณ์มหาวิทยาลัยทั้งสาขาวิชาสังคมศาสตร์และวิทยาศาสตร์ นอกจากจะให้รายละเอียดเกี่ยวกับลักษณะวิทยานิพนธ์แล้ว ยังให้รายละเอียดเกี่ยวกับกระบวนการทำวิทยานิพนธ์ด้วย สำหรับการวิเคราะห์เนื้อหาวิทยานิพนธ์ของนิสิตสาขาวิชาสังคมศาสตร์ เช่น ผลงานวิจัยของอัจฉรา ดิสวัฒน์ (2528) ให้ข้อสรุปเกี่ยวกับระยะเวลาที่ใช้สำหรับทำวิทยานิพนธ์ของนิสิตว่านิสิตส่วนใหญ่ใช้เวลามากกว่า 1 ภาคการศึกษาหรือตั้งแต่ 4 เดือนขึ้นไป

2.2 งานวิจัยที่ศึกษาตัวแปรที่ส่งผลต่อระยะเวลาที่ใช้ในการศึกษามีหลายฉบับ เช่น งานวิจัยของนันทนา รัตนอาภา (2526) และอนงค์ ปิยะกมลานนท์ (2530) งานวิจัยดังกล่าวให้ข้อสรุปที่เกี่ยวข้องกับระยะเวลาที่ทำวิทยานิพนธ์สอดคล้องกับอัจฉรา ดิสวัฒน์ ดังนี้

นันทนา รัตนอาภา (2526) เปรียบเทียบตัวแปรที่เกี่ยวข้องกับการสำเร็จการศึกษาในเวลา 2 และมากกว่า 2 ปีการศึกษาของนิสิตคณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ที่เข้าศึกษาในปีการศึกษา 2516-2520 ปรากฏว่า นิสิตทั้งกลุ่มที่สำเร็จการศึกษาในเวลา 2 และมากกว่า 2 ปีการศึกษาของทุกภาควิชาใช้เวลาในการศึกษาเนื้อหาวิชา (course work) 3 ภาคการเรียน

ส่วนเวลาที่ใช้ในการทำวิทยานิพนธ์นั้นกลุ่มที่สำเร็จการศึกษาในเวลา 2 ปีการศึกษาใช้เวลาทำวิทยานิพนธ์เสร็จภายใน 6 เดือน ขณะที่นิสิตกลุ่มสำเร็จการศึกษาใช้เวลามากกว่า 2 ปีการศึกษาใช้เวลาทำวิทยานิพนธ์ตั้งแต่ 7 เดือนขึ้นไป เมื่อพิจารณางานวิจัยของอนงค์ ปิยะกมลานนท์ (2530) ซึ่งศึกษาวิเคราะห์จำแนกกลุ่มนิสิตที่ใช้เวลาต่ำสุดและสูงสุดของหลักสูตรในการสำเร็จการศึกษาระดับมหาบัณฑิต สาขาวิชาสังคมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ที่เข้าศึกษาในปีการศึกษา 2519-2523 พบว่า กลุ่มที่ใช้เวลาต่ำสุดส่วนใหญ่ใช้เวลาศึกษาเนื้อหาวิชา 3 ภาคการศึกษา และกลุ่มที่ใช้เวลาสูงสุดส่วนใหญ่ใช้เวลา 4 ภาคการศึกษา และเมื่อพิจารณาเวลาที่ใช้ในการทำวิทยานิพนธ์ พบว่า กลุ่มที่สำเร็จการศึกษาโดยใช้เวลาต่ำสุดนั้น ทำวิทยานิพนธ์ใช้เวลา 6 เดือนและอีกกลุ่มใช้เวลามากกว่า 24 เดือนซึ่งสอดคล้องกับงานวิจัยของอัจฉรา ดิสวัฒน์ที่กล่าวมาแล้ว แสดงว่า การทำวิทยานิพนธ์เป็นสาเหตุหลักของระยะเวลาที่ใช้ในการศึกษาของนิสิตแตกต่างกัน

นอกจากนี้งานวิจัยของนันทนา รัตนอาภาและอนงค์ ปิยะกมลานนท์ ยังให้ข้อสรุปเกี่ยวกับปัจจัยจำแนกระหว่างกลุ่มนิสิตที่ใช้ระยะเวลาในการศึกษาต่างกันสอดคล้องกับงานวิจัยของพรทิพย์ ทิพย์พืช (2528) ที่ศึกษาตัวแปรที่สัมพันธ์กับระยะเวลาที่ใช้ในการศึกษาระดับปริญญาโทของข้าราชการครู กรมสามัญศึกษาซึ่งได้รับอนุมัติให้ลาศึกษาต่อเต็มเวลา ระหว่างปีการศึกษา 2521-2523 ตัวแปรที่ศึกษาสรุปได้ว่ามี 5 ตัวแปรหลัก คือ 1) ลักษณะการลาศึกษา นิสิตบางคนศึกษาโดยการลามาศึกษาเต็มเวลา แต่บางคนไม่ได้ลาศึกษา หรือลาศึกษาแต่เพียงครบเนื้อหาวิชาแล้วกลับไปทำงานควบคู่กับทำวิทยานิพนธ์ 2) ภาระรับผิดชอบในหน้าที่การงาน 3) ปัญหาการทำวิทยานิพนธ์ เช่นการค้นคว้าเอกสารและงานวิจัยที่เกี่ยวข้อง ความร่วมมือจากกลุ่มตัวอย่างและหน่วยงานที่เกี่ยวข้อง การวิเคราะห์ข้อมูลด้วยคอมพิวเตอร์ การเขียนรายงานวิจัย 4) ปัญหาส่วนตัวนิสิตด้านอารมณ์ ด้านการปรับตัวในการเรียน เช่น การใช้ห้องสมุด การวางแผนการเรียน ด้านการเงิน และ 5) ปัญหาด้านคุณลักษณะอาจารย์ที่ปรึกษาวิทยานิพนธ์ เช่น เวลาที่ให้คำปรึกษา

2.3 งานวิจัยที่ศึกษาตัวแปรที่ส่งผลต่อความสำเร็จในการทำวิทยานิพนธ์ งานวิจัยของกริสนา นกสกุล (2531) ศึกษาองค์ประกอบที่เกี่ยวข้องกับความสำเร็จในการทำวิทยานิพนธ์ของนิสิตบัณฑิตศึกษามหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร 4 ด้าน ได้แก่ องค์ประกอบด้านคุณลักษณะของอาจารย์ที่ปรึกษาวิทยานิพนธ์ ด้านคุณลักษณะของนิสิต ด้านประสบการณ์วิชาการของนิสิต และด้านการบริการวิชาการของมหาวิทยาลัย ผลปรากฏว่า คุณลักษณะของอาจารย์มีความเกี่ยวข้องกับความสำเร็จในการทำวิทยานิพนธ์มากที่สุด รองลงมา คือ

องค์ประกอบด้านคุณลักษณะของนิสิต ด้านประสบการณ์วิชาการของนิสิต และด้านการบริการวิชาการของมหาวิทยาลัย ตามลำดับ และงานวิจัยของสิริรัตน์ คุณจักร (2539) วิเคราะห์จำแนกตัวแปรที่ส่งผลต่อความสำเร็จในการทำวิทยานิพนธ์ระหว่างมหาบัณฑิตทางสังคมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ที่สำเร็จการศึกษาในเวลา 2 และมากกว่า 2 ปีการศึกษา ได้ผลสรุปว่าปัจจัยที่สามารถจำแนกการทำวิทยานิพนธ์สำเร็จระหว่างกลุ่มทั้งสองประกอบด้วยปัจจัย 2 ด้าน คือ ด้านนิสิตและด้านปัจจัยเสริม ปัจจัยด้านนิสิต ได้แก่ อายุ การลาศึกษาต่อ ลักษณะนิสัยส่วนบุคคล เช่น ลักษณะนิสัยที่เอื้อต่อการวิจัย ความรู้ความสามารถด้านวิชาการ เป็นต้น ด้านปัจจัยเสริม ได้แก่ เวลาในการให้คำปรึกษาของอาจารย์ที่ปรึกษาวิทยานิพนธ์ การรวบรวมข้อมูล ความถูกต้องและรวดเร็วของผู้พิมพ์รายงาน

3. งานวิจัยเกี่ยวกับการประยุกต์โมเดลล็อกลิเนียร์

เท่าที่ผู้วิจัยทราบยังไม่มีงานวิจัยในประเทศไทยที่วิเคราะห์ข้อมูลด้วยการวิเคราะห์ล็อกลิเนียร์ในสาขาครุศาสตร์หรือศึกษาศาสตร์ มีแต่งานวิจัยในสาขาอื่น ๆ เช่น งานวิจัยของชื่นชม เจริญยุทธ (2529) อ้างถึงในสุชาติ ประสิทธิ์รัฐสินธุ์ และกรรณิการ์ สุขเกษม (2533) ที่ศึกษาปัจจัยที่ส่งผลต่อระยะเวลาการเลี้ยงบุตรด้วยนมมารดาของสตรีครรภ์แรกที่ทำงานนอกบ้านในเขตกรุงเทพมหานคร โดยกำหนดให้ระยะเวลาการเลี้ยงบุตรด้วยนมมารดาเป็นตัวแปรตาม มี 2 ค่า และศึกษาตัวแปรต้น 8 ตัวแปร แต่ละตัวแปรมี 2 ค่า แยกวิเคราะห์ตัวแปรที่ส่งผลต่อตัวแปรตามครั้งละ 2 ตัวแปร ด้วยตารางการณ์จร 3 มิติ

ส่วนการประยุกต์โมเดลล็อกลิเนียร์ในการศึกษาความสัมพันธ์เชิงสาเหตุที่ใช้เทคนิคการวิเคราะห์อิทธิพล (path analysis) ในประเทศไทยเท่าที่ผู้วิจัยทราบยังไม่มี

4. สมมติฐานงานวิจัย

จากรายงานประจำปี จุฬาลงกรณ์มหาวิทยาลัย ปีการศึกษา 2536 สรุปได้ว่านิสิตที่สำเร็จการศึกษาในเวลา 2 ปีการศึกษามีจำนวนน้อยกว่านิสิตที่สำเร็จการศึกษาในเวลา 3-4 ปีการศึกษา สาเหตุสำคัญคือ นิสิตอาจมีปัญหาในการทำวิทยานิพนธ์แตกต่างกันจึงไม่สามารถเสนอวิทยานิพนธ์ได้ทันหรือไม่สามารถจัดทำวิทยานิพนธ์เสร็จทันในเวลาที่กำหนดไว้ ปัญหาการทำวิทยานิพนธ์เริ่มมีตั้งแต่การพัฒนาโครงการเสนอวิทยานิพนธ์จนกระทั่งถึงกระบวนการในการทำวิทยานิพนธ์ สำหรับกระบวนการในการพัฒนาโครงการเสนอวิทยานิพนธ์นั้น ชนะ กองไตรย์ (2537) ได้ศึกษาไว้ว่าประกอบด้วย 4 ขั้นตอน คือ ขั้นตอนที่ 1 การเริ่มต้นของกระบวนการพัฒนาโครงการเสนอวิทยานิพนธ์ ขั้นตอนที่สอง การได้อาจารย์ที่ปรึกษา

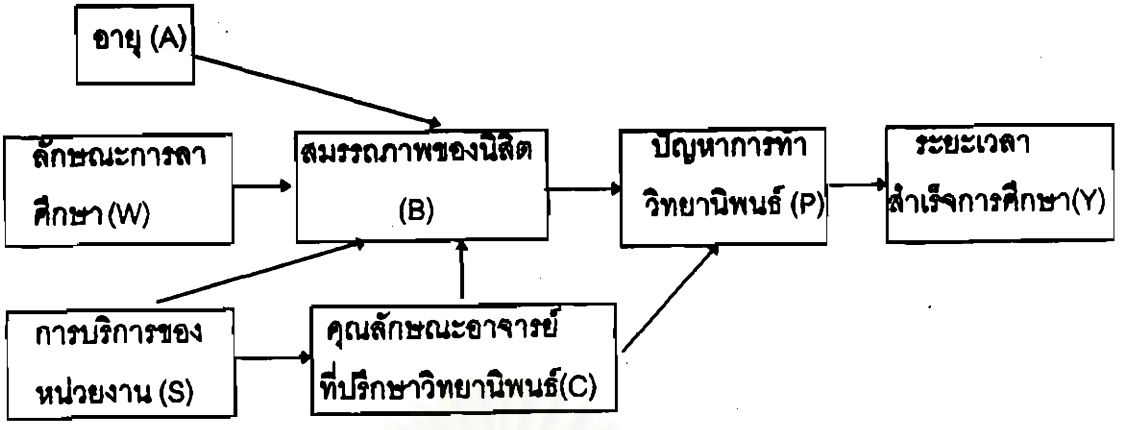
วิทยานิพนธ์ ขั้นตอนที่สาม การเขียนโครงการเสนอวิทยานิพนธ์ และขั้นตอนที่สี่ การเสนอโครงการเสนอวิทยานิพนธ์เพื่อให้ภาควิชาพิจารณา

นอกจากข้อสรุปดังกล่าวแล้ว ชนะ กองไตรย์ (2537) ให้ข้อสรุปเกี่ยวกับปัจจัยที่ส่งเสริมและอุปสรรคที่ทำให้นิสิตมีกระบวนการพัฒนาโครงการเสนอวิทยานิพนธ์แตกต่างกัน สิ่งที่ส่งเสริมหรือสนับสนุน ได้แก่ นโยบายของภาควิชา อาจารย์ที่ปรึกษามีบารมี หรืออำนาจในการต่อรอง การเอาใจใส่ต่อนิสิต นิสิตมีความรู้ด้านระเบียบวิธีวิจัย มีแหล่งค้นคว้าที่พอเพียง ส่วนอุปสรรค ได้แก่ นิสิตขาดความรู้ด้านระเบียบวิธีและการวิเคราะห์ข้อมูล นิสิตขาดการติดต่อกับอาจารย์ที่ปรึกษาวิทยานิพนธ์ และนิสิตไม่มีเวลาในการพัฒนาโครงการเสนอวิทยานิพนธ์ ปัจจัยเหล่านี้สอดคล้องกับผลงานวิจัยของนักวิจัยที่ได้กล่าวแล้วเกี่ยวกับการศึกษาปัจจัยที่ส่งผลต่อระยะเวลาที่ใช้ในการศึกษาต่างกัน และปัจจัยที่มีผลต่อความสำเร็จในการทำวิทยานิพนธ์ ปัจจัยเหล่านี้สรุปได้ว่าระยะเวลาที่ใช้ในการศึกษาต่างกันและความสำเร็จในการทำวิทยานิพนธ์มีสาเหตุมาจากตัวนิสิตเอง และนิสิตแต่ละคนจะมีปัจจัยเสริมและอุปสรรคแตกต่างกัน

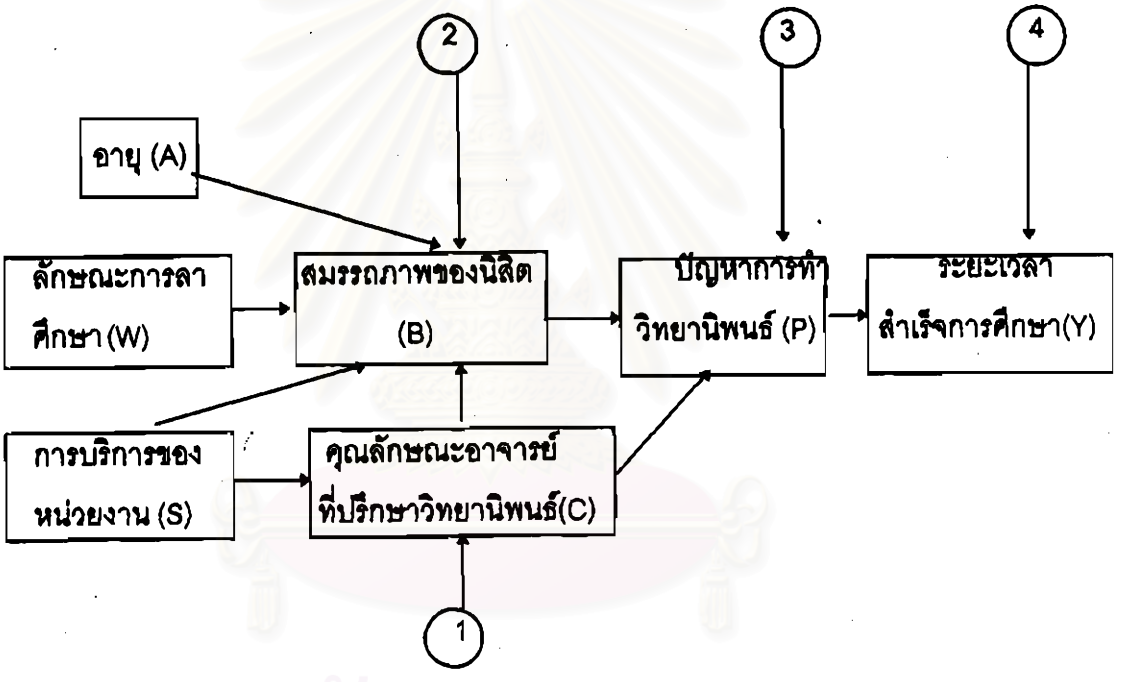
งานวิจัยนี้ผู้วิจัยศึกษาความสัมพันธ์เชิงสาเหตุของปัจจัยที่ส่งผลต่อความสำเร็จในการทำวิทยานิพนธ์ของมหาบัณฑิตทางสังคมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย โดยประยุกต์ใช้กับโมเดลล็อกลิเนียร์ และกำหนดให้ระยะเวลาที่ใช้ในการศึกษา (Y) เป็นตัวแปรตาม ผู้วิจัยคัดเลือกปัจจัยที่เป็นเหตุจากผลสรุปที่สอดคล้องกันของงานวิจัยที่กล่าวแล้ว ได้แก่ 1) สภาพทั่ว ๆ ไปของนิสิต ประกอบด้วย อายุ (A) ลักษณะการลาศึกษา (W) และสมรรถภาพการทำวิทยานิพนธ์ (B) และ 2) ปัจจัยเสริม ได้แก่ คุณลักษณะของอาจารย์ที่ปรึกษาวิทยานิพนธ์(C) การบริการของหน่วยงาน (S) และปัญหาการทำวิทยานิพนธ์ (P) โดยมีแนวคิดในการสร้างโมเดลสมมติฐานการวิจัยว่า นิสิตระดับปริญญาโทจะสำเร็จการศึกษาได้ต้องได้รับคะแนนเฉลี่ยสะสมไม่ต่ำกว่า 3.00 และวิทยานิพนธ์ได้รับอนุมัติให้เป็นส่วนหนึ่งของหลักสูตร และการทำวิทยานิพนธ์ถือว่าเป็นประสบการณ์ครั้งแรกในการทำวิจัยของนิสิต ฉะนั้น นิสิตทุกคนจะต้องพบกับปัญหาในการทำวิทยานิพนธ์มากมาย นิสิตแต่ละคนจะแก้ปัญหาได้ต่างกันทำให้นิสิตสำเร็จการศึกษาในเวลาต่างกัน และเนื่องจากความแตกต่างของสมรรถภาพของนิสิตและปัจจัยเสริมตลอดจนอุปสรรคในการทำวิทยานิพนธ์จึงส่งผลให้นิสิตแก้ปัญหาได้แตกต่างกันด้วย และจากการวิเคราะห์ตัวประกอบของจารึก อาจวารินทร์ (2529) พบว่า นักวิจัยควรมีสมรรถภาพทั้งทางด้านความรู้ความสามารถในระเบียบวิธีวิจัยและสถิติวิเคราะห์ข้อมูล ทัศนคติเชิงวิทยาศาสตร์และลักษณะนิสัยที่เอื้อต่อการวิจัย ดังนั้นนิสิตควรมีความรู้ความสามารถหรือสมรรถภาพในด้านต่าง ๆ เพื่อที่

จะประสบความสำเร็จในการทำงานอย่างมีขวัญและกำลังใจ (สิริรัตน์ คุณจักร, 2539) สำหรับงานวิจัยครั้งนี้สมรรถภาพของนิสิตจะอยู่ในระดับใดขึ้นกับปัจจัยต่าง ๆ ได้แก่ 1) คุณลักษณะนิสิตทั่ว ๆ ไป เช่น อายุ ลักษณะการลาศึกษาต่อ และสมรรถภาพของนิสิต นิสิตที่อายุต่างกันอาจส่งผลต่อสมรรถภาพของนิสิตได้ และนิสิตที่มาเรียนโดยลาศึกษาต่ออาจมีผลดีต่อสมรรถภาพการทำวิทยานิพนธ์มากกว่านิสิตที่ไม่ได้ลาศึกษา 2) คุณลักษณะของอาจารย์ที่ปรึกษาวิทยานิพนธ์ เพราะอาจารย์ที่ปรึกษาวิทยานิพนธ์เป็นบุคคลสำคัญต่อนิสิตอย่างมาก การได้รับคำแนะนำที่ดี มีคุณภาพ ให้แนวคิดใหม่ ๆ แก่นิสิตทำให้นิสิตเกิดการเรียนรู้ในการทำวิทยานิพนธ์ดำเนินไปได้ตามจุดมุ่งหมายได้ ฉะนั้น ถ้าหากนิสิตคนใดมีโอกาสได้รับเวลาและคำแนะนำที่มีคุณภาพจากอาจารย์ที่ปรึกษาวิทยานิพนธ์ก็มีผลดีต่อสมรรถภาพของนิสิตซึ่งจะช่วยส่งผลให้นิสิตสามารถแก้ปัญหาที่เกิดขึ้นขณะทำวิทยานิพนธ์ และประสบความสำเร็จในการทำวิทยานิพนธ์ได้เช่นกัน โดยเฉพาะถ้านิสิตรู้สึกว่าได้รับโอกาสนี้จากอาจารย์ในระดับดี ๆ ก็ยิ่งส่งผลให้นิสิตมีสมรรถภาพมากขึ้น แก้ปัญหาได้เร็วขึ้น และประสบความสำเร็จในการทำวิทยานิพนธ์ได้เร็วขึ้น 3) การบริการของสถาบัน ในที่นี้หมายถึงการอำนวยความสะดวกของบัณฑิตวิทยาลัย บัณฑิตคณะ และภาควิชา ในด้านบริการห้องสมุด เพราะห้องสมุดเป็นแหล่งค้นคว้าหาความรู้ที่สำคัญยิ่งกับนิสิตและอาจารย์ที่ปรึกษา ดังนั้น ตำรา หนังสือ เอกสารวิชาการต่าง ๆ ในห้องสมุดควรมีอย่างพอเพียง ตลอดจนอุปกรณ์สื่อสารในการค้นคว้าหาความรู้ก็ควรทันสมัยและทันต่อการเปลี่ยนแปลงของสังคมโลกด้วย ฉะนั้น สิ่งเหล่านี้สถาบันควรจัดหาบริการให้พอเพียงซึ่งจะเป็นการเอื้อให้นิสิตประสบความสำเร็จในการทำวิทยานิพนธ์ได้

อย่างไรก็ตาม ความสัมพันธ์ของปัจจัยต่างๆ ที่กล่าวข้างต้นมีได้ทั้งในเทอมของอิทธิพลหลักและปฏิสัมพันธ์ แต่เนื่องจากยังไม่มีรายงานการวิจัยหรือทฤษฎีที่ระบุถึงอิทธิพลของเทอมปฏิสัมพันธ์ระหว่างตัวแปรเหตุที่มีต่อความสำเร็จในการทำวิทยานิพนธ์และส่งผลให้นิสิตใช้ระยะเวลาในการศึกษาต่างกัน ดังนั้น จากแนวคิดข้างต้น ผู้วิจัยจึงกำหนดโมเดลความสำเร็จในการทำวิทยานิพนธ์ 2 แบบ แบบแรก โมเดลประกอบด้วยเฉพาะอิทธิพลหลักของตัวแปรเหตุที่มีอิทธิพลต่อตัวแปรตาม ดังรูปภาพที่ 3 ส่วนแบบที่ 2 ประกอบด้วยอิทธิพลหลักและปฏิสัมพันธ์อันดับหนึ่งของตัวแปรเหตุที่มีอิทธิพลต่อตัวแปรตาม ดังรูปภาพที่ 4



รูปภาพที่ 3 โมเดลสมมติฐาน แบบ 1



รูปภาพที่ 4 โมเดลสมมติฐาน แบบ 2

หมายเหตุ

(1) ปฏิสัมพันธ์อันดับหนึ่ง AW, AS และ WS

(2) ปฏิสัมพันธ์อันดับหนึ่ง AW, AS, AC, WS, WC และ SC

(3) ปฏิสัมพันธ์อันดับหนึ่ง AW, AS, AC, AB, WS, WC, WB, SC, SB และ CB

(4) ปฏิสัมพันธ์อันดับหนึ่ง AW, AS, AC, AB, AP, WS, WC, WB, WP, SC, SB, SP, CB, CP และ BP