บทที่ 2

วรรณคดีที่เกี่ยวข้อง

ผู้วิจัยจะนำเสนอผลการศึกษาค้นคว้าเอกสารและงานวิจัยที่เกี่ยวข้องโดยแบ่งการนำเสนอออกเป็น 4 ตอน ดังนี้คือ ตอนที่ 1 ความรู้ทั่วไปเกี่ยวกับโมเดลล็อกลิเนียร์ และขั้นตอนในการวิเคราะห์ล็อกลิเนียร์สำหรับตัวแปรมาตรนามบัญญัติ โดยกล่าวถึงมโนทัศน์ (concept) ของการวิเคราะห์ล็อกลิเนียร์ การสร้างตารางการณ์จรและการคำนวณความถี่ที่คาดหวัง การสร้างโมเดลล็อกลิเนียร์ การประมาณค่าพารามิเตอร์ การคัดเลือกโมเดลล็อกลิเนียร์ การตรวจสอบความเหมาะสมของโมเดล และการประมาณค่าพารามิเตอร์ลอการิทึมของอัตราส่วนแต้มต่อ ตอนที่ 2 การวิเคราะห์ล็อกลิเนียร์มาตรอันดับ กล่าวถึงมโนทัศน์ของการวิเคราะห์ล็อกลิเนียร์สำหรับตัวแปรมาตรอันดับ การทดสอบความสอดคล้องกลมกลืนของโมเดล การทดสอบเงื่อนไขของความเป็นอิสระ การคำนวณค่าอัตราส่วนแต้มต่อ และโมเดลโลจิทสำหรับการวิเคราะห์ล็อกลิเนียร์มาตรอันดับ ตอนที่ 3 งานวิจัยที่เกี่ยวข้องกับการศึกษาถึงตัวแปรที่สัมพันธ์กับการศึกษาต่อและไม่ศึกษาต่อของนักเรียน ในตอนนี้ผู้วิจัยนำเสนอผลการศึกษาค้นคว้าและสังเคราะห์งานวิจัยที่เกี่ยวข้องกับตัวแปรที่สัมพันธ์กับการศึกษาต่อและไม่ศึกษาต่อของนักเรียนเพื่อเป็นการคัดเลือกและกำหนดตัวแปรในโมเดลความคาดหวังในการศึกษาต่อของนักเรียนและกำหนดสมมุติฐาน และตอนที่ 4 กล่าวถึงการนำการวิเคราะห์ล็อกลิเนียร์มาใช้ในประเทศไทย

ตอนที่ 1 ความรู้ทั่วไปเกี่ยวกับโมเดลล็อกลิเนียร์

การวิเคราะห์ข้อมูลที่เป็นข้อมูลประเภทกลุ่ม (categorical data) สามารถทำได้โดยการนำข้อมูลใส่ลงไปในตารางการณ์จร (contingency table) แต่การวิเคราะห์ข้อมูลประเภทนี้มีข้อจำกัดในเรื่องของทฤษฎีและเทคนิคในการคำนวณ การวิเคราะห์ข้อมูลจึงทำได้เฉพาะตาราง 2 มิติ และจากการพัฒนาของ Karl Pearson (1900) ได้พัฒนาการวิเคราะห์ข้อมูลโดยใช้การทดสอบความกลมกลืนแบบไค-สแควร์ (χ^2 goodness of fit) ต่อมาได้มีนักสถิติพยายามพัฒนาการวิเคราะห์ข้อมูลมากกว่า 2 มิติ พร้อมๆกัน เช่น Birch Bichop, Fienberg Goodman, Grizzle Williams และ Haberman (ทวีพร บุญวาณิช, 2541; Kennedy, 1983) จนได้เทคนิควิธีการวิเคราะห์หลายๆ รูปแบบที่นำมาใช้กับการวิเคราะห์ข้อมูลพร้อมๆ กันจากตารางการณ์จรพหุมิติ (multidimensional contingency table) จากการเสนอผลงานของนักสถิติในช่วงเวลาต่างๆ จึงกลายเป็นเทคนิควิธีการ

วิเคราะห์ที่เรียกว่า การวิเคราะห์ตารางการณ์จรแบบล็อกลิเนียร์ (log-linear contingency table analysis) หรือ เรียกกันว่า การวิเคราะห์ล็อกลิเนียร์ (log-linear analysis) และจากการที่โมเดลล็อกลิเนียร์เป็นโมเดลแบบคูณ (multiplicative model) จึงต้องคำนวณด้วยค่าลอการิทึม (logarithm) ทำให้โมเดลแบบคูณอยู่ในรูปของโมเดลแบบบวก การเปลี่ยนรูปโมเดลนี้จึงเป็นที่มาของโมเดลล็อกลิเนียร์ (log-linear model) (Kennedy, 1983; Steven 1996)

การวิเคราะห์ล็อกลิเนียร์เป็นผลมาจากการบูรณาการเทคนิควิธีการวิเคราะห์ข้อมูล 3 แบบด้วยกัน คือ การวิเคราะห์ความแปรปรวน (analysis of variance) การวิเคราะห์การถดถอย (regression analysis) และการทดสอบความกลมกลืนแบบไค-สแควร์ หรือการทดสอบภาวะสารูปสนิท (chi-square goodness of fit) โดยการวิเคราะห์ล็อกลิเนียร์จะมีลักษณะคล้ายกับการวิเคราะห์ความแปรปรวนตรงที่สามารถทำให้ทราบอิทธิพลหลัก (main effect) และอิทธิพลปฏิสัมพันธ์ (interaction effect) ของตัวแปร และยังมีโมเดลในการนำเสนอความสัมพันธ์ระหว่างตัวแปรที่ศึกษา (Kennedy, 1983) แต่การวิเคราะห์ล็อกลิเนียร์ยังมีข้อแตกต่างจากการวิเคราะห์ความแปรปรวนสรุปได้ดังนี้ คือ (ทวีพร บุญวาณิช, 2541; Kennedy, 1983; Steven, 1996)

1. โมเดลในการทำนายของการวิเคราะห์ความแปรปรวนจะทำนายค่าเฉลี่ยของตัวแปรตามด้วยอิทธิพลของตัวแปรอิสระ และเป็นโมเดลแบบบวก (additive model) ส่วนโมเดลการวิเคราะห์ล็อกลิเนียร์ เป็นโมเดลที่ทำนายโอกาสในการเกิดค่าความถี่ที่คาดหวัง (expected frequency) ของตารางการณ์จรของตัวแปรตาม ด้วยชุดอิทธิพลของตัวแปรอิสระ และเป็นโมเดลแบบคูณ (multiplicative model)

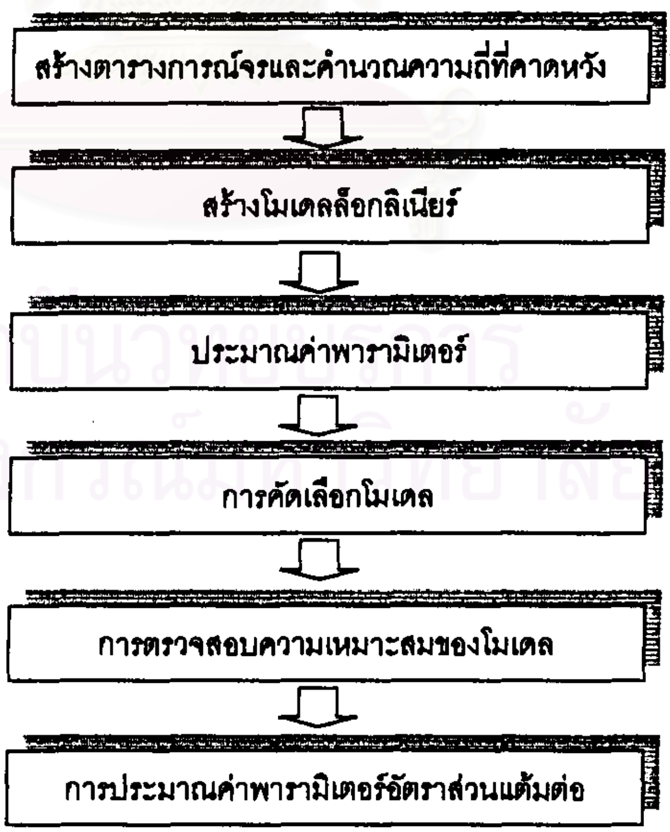
2. ในการวิเคราะห์ความแปรปรวน นักวิจัยสามารถสร้างโมเดลได้เพียงโมเดลเดียวเท่านั้น แต่ในการวิเคราะห์ล็อกลิเนียร์ นักวิจัยสามารถสร้างโมเดลได้หลายโมเดลเพื่อใช้ในการพิจารณาต่อไป

3. การวิเคราะห์ความแปรปรวน เป็นการศึกษาความสัมพันธ์ระหว่างตัวแปรโดยกำหนดให้ตัวแปรตัวหนึ่งเป็นตัวแปรตาม และศึกษาถึงอิทธิพลของชุดของตัวแปรอิสระที่มีต่อตัวแปรตาม ส่วนการวิเคราะห์ล็อกลิเนียร์ สามารถวิเคราะห์ความสัมพันธ์ได้ 2 ลักษณะ คือ ลักษณะที่หนึ่ง การวิเคราะห์ความสัมพันธ์แบบสมมาตร (symmetrical relationship) เป็นการศึกษาความสัมพันธ์ระหว่างตัวแปรสองตัวโดยไม่ระบุว่าตัวแปรใดเป็นตัวแปรตามและตัวแปรใดเป็นตัวแปรอิสระ ลักษณะที่สองคือ การวิเคราะห์ความสัมพันธ์แบบอสมมาตร (asymmetrical relationship)

เป็นการศึกษาความสัมพันธ์เมื่อกำหนดให้ตัวแปรหนึ่งเป็นตัวแปรตาม จากลักษณะดังกล่าว จะเห็นได้ว่า การวิเคราะห์ความแปรปรวนเป็นการวิเคราะห์ความสัมพันธ์แบบอสมมาตรเท่านั้น

4. ในการวิเคราะห์ความแปรปรวนจะกำหนดให้ตัวแปรตามเป็นตัวแปรต่อเนื่อง (continuous variables) ตัวแปรอิสระเป็นมาตรนามบัญญัติ แต่ในการวิเคราะห์ล็อกลิเนียร์กำหนดให้ตัวแปรอิสระและตัวแปรตามเป็นตัวแปรกลุ่ม (categorical variables) โดยอาจจะเป็นตัวแปรมาตรนามบัญญัติ หรือมาตรอันดับก็ได้

การวิเคราะห์ล็อกลิเนียร์สำหรับตัวแปรมาตรนามบัญญัติและตัวแปรมาตรอันดับ มีกระบวนการในการวิเคราะห์เหมือนกันทั้งแบบสมมาตร และแบบอสมมาตร โดยกระบวนการวิเคราะห์ได้แบ่งออกเป็น 6 ขั้นตอนด้วยกัน คือ ขั้นตอนแรก สร้างตารางการณ์จรและคำนวณความถี่ที่คาดหวัง ขั้นตอนที่สอง สร้างโมเดลล็อกลิเนียร์ (log-linear modeling) ขั้นตอนที่สาม ประมาณค่าพารามิเตอร์ (parameter estimates) ขั้นตอนที่สี่ คัดเลือกโมเดล (model selection) ด้วยการทดสอบความกลมกลืน (fitting log-linear model) ขั้นตอนที่ห้า ตรวจสอบความเหมาะสมของโมเดล และขั้นตอนที่หก ประมาณค่าพารามิเตอร์อัตราส่วนแต้มต่อ ดังภาพที่ 1 โดยมีรายละเอียดแต่ละหัวข้อต่อไปนี้ (ทวีพร บุญวาณิช, 2541; Kennedy, 1983; Steven, 1996)



ภาพที่ 1 กระบวนการในการวิเคราะห์ล็อกลิเนียร์

1.การสร้างตารางการณ์จร และการคำนวณความถี่ที่คาดหวัง

1.1 การสร้างตารางการณ์จร (Creating Contingency Table)

การสร้างตารางการณ์จรของการวิเคราะห์ล็อกลิเนียร์นั้น ข้อมูลในแต่ละเซลล์ของตาราง คือความถี่ (frequency) ที่ได้มาจากการแจงนับหน่วยตัวอย่างที่สุ่มจากประชากรตามข้อตกลงว่า หน่วยตัวอย่างแต่ละหน่วยต้องถูกจำแนกลงในแต่ละเซลล์ได้อย่างอิสระ (independent) เนื่องจากลักษณะของตัวแปรแบ่งเป็นกลุ่มๆ ที่ไม่ต่อเนื่อง การแจกแจงความถี่ของตารางการณ์จรในแต่ละเซลล์นี้เป็นแบบไม่เกิดร่วม (mutually exclusive) และแต่ละเซลล์ต้องมีความถี่มากพอที่จะวิเคราะห์ ตารางจะมีขนาด (dimension) หรือจำนวนเซลล์เป็นเท่าไร ขึ้นอยู่กับจำนวนตัวแปรและจำนวนค่าสังเกตของตัวแปรนั้นๆ เช่นการศึกษาตัวแปร 2 ตัวแปร คือตัวแปร A และ ตัวแปร B ตัวแปร A มี i ค่า ส่วนตัวแปร B มี j ค่า ในการสร้างตารางการณ์จรจึงได้ตารางขนาด 2 มิติ ที่มีขนาด i X j โดยกำหนดให้ i เป็นแถว (rows) และ j เป็นสดมภ์ (columns) จะได้เซลล์ ij เกิดขึ้น ในการดูถึงอิทธิพลหลักของ A จะพิจารณาจากค่าความถี่ที่เป็นผลรวมในแต่ละแถว และอิทธิพลหลักของ B จะพิจารณาจากความถี่ในแต่ละสดมภ์ ถ้าเป็นการศึกษาตัวแปร 3 ตัว ได้แก่ A B C โดยที่ A มี i ค่า B มี j ค่า และ C มี k ค่า จะได้ตารางขนาด 3 มิติ จึงสรุปได้ว่าตารางจะมีขนาดกี่มิติขึ้นอยู่กับจำนวนตัวแปร ตารางเหล่านี้จะวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรแบบสมมาตรและแบบอสมมาตรได้ทั้งสองวิธี โดยกำหนดสัญลักษณ์ที่ใช้ดังต่อไปนี้

i	แทน	แถวในตาราง หรือระดับค่าของตัวแปร A
j	แทน	สดมภ์ในตาราง หรือระดับค่าของตัวแปร B
f_{ij}	แทน	ความถี่ที่สังเกตได้ (observed frequency) ในแถวที่ i สดมที่ j หรือความถี่ที่สังเกตได้ เซลล์ ij (cell frequency)
$n_{i.}$	แทน	ผลรวมความถี่ที่สังเกตได้ (marginal observed frequency) ตามแถวที่ $i = \Sigma f_{ij}$
$n_{.j}$	แทน	ผลรวมความถี่ที่สังเกตได้ (marginal observed frequency) ตามสดมภ์ที่ $j = \Sigma f_{ij}$
n	แทน	ผลรวมความถี่ที่สังเกตได้ทั้งหมด $= \Sigma\Sigma f_{ij}$

1.2 การคำนวณค่าความถี่ที่คาดหวัง (Expected Frequency)

เนื่องจากการวิเคราะห์ล็อกลิเนียร์เป็นการทำนายโอกาสในการเกิดความถี่ที่คาดหวัง จึงต้องมีการคำนวณค่าความถี่ที่คาดหวังในแต่ละเซลล์ การวิเคราะห์ล็อกลิเนียร์มีการคำนวณค่าความถี่ที่คาดหวังคล้ายกับการคำนวณค่าความถี่ที่คาดหวังของการนำไปหาค่าสถิติไค-สแควร์ เพียงแต่การคำนวณความถี่ที่คาดหวังในการวิเคราะห์ล็อกลิเนียร์มีข้อแตกต่างกันหลายวิธีตาม ข้อกำหนดในการสร้างโมเดล ได้แก่ การคำนวณค่าความถี่ที่คาดหวัง กรณีไม่มีอิทธิพลใดๆ จากตัวแปร A และ B กรณีได้รับอิทธิพลหลักจากตัวแปรใดตัวแปรหนึ่ง กรณีได้รับทั้งอิทธิพลหลักของตัวแปร A และอิทธิพลหลักของตัวแปร B และกรณีได้รับทั้งอิทธิพลหลักของตัวแปร A อิทธิพลหลักของตัวแปร B และปฏิสัมพันธ์ระหว่าง A กับ B ดังตัวอย่างต่อไปนี้

กำหนดให้ตารางการณ์จรประกอบด้วยตัวแปรจำนวน 2 ตัวแปร คือตัวแปร A และ B ในตารางขนาด $i \times j$ เมื่อ $i = 2$ และ $j = 3$ จะได้ตารางการณ์จร 2 แถว 3 สดมภ์ รวม 6 เซลล์ เมื่อ

p_{ij} = ความน่าจะเป็นในการจำแนกหน่วยตัวอย่างลงแถวที่ i สดมภ์ที่ j

n = จำนวนหน่วยตัวอย่าง หรือขนาดของกลุ่มตัวอย่างในการวิเคราะห์

F_{ij} = ความถี่ที่คาดหวังในแถวที่ i สดมภ์ที่ j

$= np_{ij}$

ซึ่งจะสามารถคำนวณค่าความถี่ที่คาดหวังในแต่ละเซลล์ได้ 4 ลักษณะ โดยมีวิธีการคำนวณดังนี้

ลักษณะที่ 1 การคำนวณความถี่ที่คาดหวัง กรณีไม่มีอิทธิพลใด ๆ จากตัวแปร A และ B

ค่าความถี่ที่คาดหวังในแต่ละเซลล์ คือ จำนวนหน่วยตัวอย่างที่มีความน่าจะเป็นที่จะถูกจำแนกลงในแถวที่ i สดมภ์ที่ j โดยไม่ได้รับอิทธิพลใดๆ จากตัวแปร A และ Bเลย หมายความว่า โอกาสที่หน่วยตัวอย่างถูกจำแนกลงเซลล์ ij นั้นมีค่าเท่ากัน จึงทำให้ความถี่ที่คาดหวังในเซลล์ ij มีค่าเท่า ๆ กัน ในกรณีนี้ตารางการณ์จรมี 6 เซลล์ และจำนวนหน่วยตัวอย่างมี n หน่วย ผลการคำนวณค่าความถี่ที่คาดหวังแสดงได้ดังตารางที่ 1

โดยที่ $p_{ij} = 1/6$ และ $F_{ij} = n/6$

ตารางที่ 1 การแจกแจงความถี่ที่คาดหวังลงตารางการณ์จร 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 หรือตัวแปร B ตัวใดตัวหนึ่ง เพียงตัวแปรเดียว

ค่าความถี่ที่คาดหวังในแต่ละเซลล์ คือ จำนวนหน่วยตัวอย่างที่มีความน่าจะเป็นที่จะถูกจำแนกลงในแถวที่ i สดมภ์ที่ j โดยได้รับอิทธิพลหลักจากตัวแปร A เพียงตัวเดียว ผลการคำนวณค่าความถี่ที่คาดหวังดังแสดงในตารางที่ 2 กรณีนี้ตารางการณ์จร มี 6 เซลล์ ตัวแปร A มี 2 ค่า เมื่อมีอิทธิพลจากตัวแปร A ผลรวมความถี่ที่คาดหวังในแถวที่ 1 และ 2 จึงแตกต่างกัน ทำให้ความถี่ที่คาดหวังในแต่ละสดมภ์มีค่าแตกต่างกันไปด้วย แต่ค่าความถี่ที่คาดหวังทั้ง 3 สดมภ์ในแต่ละแถวต้องเท่ากัน หมายความว่า โอกาสที่หน่วยตัวอย่างถูกจำแนกลงแต่ละสดมภ์ j แถว i มีค่าเท่า ๆ กัน โอกาสหรือค่าความน่าจะเป็นในที่นี้คือ ความน่าจะเป็นมีเงื่อนไข (conditional probability=p_{ji})นั่นเอง เนื่องจากตารางการณ์จรในแต่ละแถวมี 3 สดมภ์ และจำนวนหน่วยตัวอย่างในแถวที่ i หรือผลรวมความถี่แถวที่ i มีค่าเท่ากับ $n_{i.}$ เมื่อความถี่สดมภ์ j ในแต่ละแถว i มีค่าเท่ากัน ค่า p_{ji} = 1/3 จึงคำนวณหาความถี่ที่คาดหวังได้จากความน่าจะเป็นอย่างมีเงื่อนไข ดังนี้

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

ตารางที่ 2 การแจกแจงความถี่ที่คาดหวังลงตารางการณ์จร 2 มิติ ขนาด 2 X 3 ลักษณะ 2 กรณีได้รับอิทธิพลหลักจากตัวแปร A

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

ในกรณีที่ได้รับอิทธิพลหลักจากตัวแปร B เพียงตัวแปรเดียว ผลการคำนวณค่าความถี่ที่คาดหวังดังแสดงในตารางที่ 3 กรณีนี้ตารางการณ์จร มี 6 เซลล์ ตัวแปร B มี 3 ค่า เมื่อมีอิทธิพลจากตัวแปร B ผลรวมความถี่ที่คาดหวังในสดมภ์ที่ 1, 2 และ 3 จึงแตกต่างกัน ทำให้ความถี่ที่คาดหวังในแต่ละแถวมีค่าแตกต่างกันไปด้วย แต่ค่าความถี่ที่คาดหวังทั้ง 2 แถวในแต่ละสดมภ์ต้องเท่ากัน หมายความว่า โอกาสที่หน่วยตัวอย่างถูกจำแนกลงแต่ละแถว i สดมภ์ j มีค่าเท่า ๆ กัน โอกาสหรือค่าความน่าจะเป็นในที่นี้คือ ความน่าจะเป็นมีเงื่อนไข (conditional probability=$p_{i|j}$)นั่นเอง เนื่องจากตารางการณ์จรในแต่ละสดมภ์มี 2 แถว และจำนวนหน่วยตัวอย่างในสดมภ์ที่ j หรือผลรวมความถี่สดมภ์ที่ j มีค่าเท่ากับ $n_{.j}$ เมื่อความถี่แถว i ในแต่ละสดมภ์ j มีค่าเท่ากัน ค่า $p_{i|j}$ = 1/2 จึงคำนวณหาความถี่ที่คาดหวังได้จากความน่าจะเป็นอย่างมีเงื่อนไข ดังนี้

$$F_{ij} = n_{.j}\, p_{i|j} = n_{.j} / 2$$

ตารางที่ 3 แสดงการแจกแจงความถี่ที่คาดหวังลงตารางการณ์จร 2 มิติ ขนาด 2 X 3 ลักษณะ 2 กรณีได้รับอิทธิพลหลักจากตัวแปร B

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

ลักษณะที่ 3 การคำนวณความถี่ที่คาดหวัง กรณีได้รับทั้งอิทธิพลหลักของตัวแปร A และอิทธิพลหลักของตัวแปร B

ค่าความถี่ที่คาดหวังในแต่ละเซลล์ คือ จำนวนหน่วยตัวอย่างที่มีความน่าจะเป็นที่ถูกจำแนกลงในแถวที่ i สดมภ์ที่ j โดยได้รับอิทธิพลทั้งอิทธิพลหลักของตัวแปร A และ B ดังนั้นค่าความถี่ที่คาดหวังในแต่ละเซลล์จึงแตกต่างกันทั้งหมด ความน่าจะเป็นที่หน่วยตัวอย่างแต่ละหน่วยจะถูกจำแนกลงในแถวที่ i สดมภ์ที่ j (p_{ij}) มีค่าเท่ากับผลคูณของความน่าจะเป็นที่หน่วยตัวอย่างถูกจำแนกลงในแถวที่ i ($p_{i.}$) กับความน่าจะเป็นที่หน่วยตัวอย่างถูกจำแนกลงในสดมภ์ที่ j ($p_{.j}$) นั่นคือ

$p_{ij} = p_{i.}\, p_{.j}$ โดยที่ตัวแปร 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 และความน่าจะเป็นดังแสดงในตารางที่ 4 จะเห็นได้ว่าค่าความถี่ที่คาดหวังในตารางการณ์จรมีลักษณะการคำนวณความถี่ที่คาดหวัง ซึ่งใช้ในการวิเคราะห์ความสัมพันธ์ หรือทดสอบความเป็นอิสระ (test of independence) ระหว่างตัวแปร 2 ตัว โดยใช้สถิติไค-สแควร์ นั่นเอง

ตารางที่ 4 การแจกแจงความถี่ที่คาดหวังลงตารางการณ์จร 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 และอิทธิพลปฏิสัมพันธ์ระหว่าง A กับ B

ค่าความถี่ที่คาดหวังในแต่ละเซลล์ คือ จำนวนหน่วยตัวอย่างที่มีความน่าจะเป็นที่ถูกจำแนกลงในแถวที่ i สดมภ์ที่ j โดยได้รับอิทธิพลหลักของตัวแปร A อิทธิพลหลักของตัวแปร B และอิทธิพลปฏิสัมพันธ์ของ A และ B ค่าความถี่ที่คาดหวังในแต่ละเซลล์มีค่าเท่ากับความถี่ที่สังเกตได้ ดังตารางที่ 5

ตารางที่ 5 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

หากพิจารณาถึงการคำนวณค่าความถี่ที่คาดหวังทั้ง 4 ลักษณะดังได้กล่าวมาแล้ว จะทำให้เห็นถึงอิทธิพลหลักและอิทธิพลปฏิสัมพันธ์ จากตัวแปร A และ B ว่ามีลักษณะคล้ายคลึงกับโมเดลในการวิเคราะห์ความแปรปรวนแบบสองทาง (two way anova) นั่นเอง กล่าวคือ

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

เมื่อ Y_{ij} = คะแนนเฉลี่ยเซลล์ ij $\qquad$ F_{ij} = ความถี่ที่คาดหวังในเซลล์ ij

$\overline{Y}..$ = ค่าเฉลี่ยรวม = $\sum Y_{ij}/n$ $\qquad$ τ = ค่าเฉลี่ยรวมของความถี่ที่คาดหวัง

a_i = อิทธิพลหลักจากตัวแปร A $\qquad$ τ_i^a = อิทธิพลหลักจากตัวแปร A

b_j = อิทธิพลหลักจากตัวแปร B $\qquad$ τ_j^b = อิทธิพลหลักจากตัวแปร B

ab_{ij} = ปฏิสัมพันธ์ระหว่างตัวแปร A กับ B $\qquad$ τ_{ij}^{ab} = ปฏิสัมพันธ์ A และ B

ซึ่งในการวิเคราะห์ล็อกลิเนียร์นั้นจะนำสมการทั้ง 4 นี้ไปใช้

2. การสร้างโมเดลล็อกลิเนียร์ (Log-linear modeling)

จากสมการการวิเคราะห์ล็อกลิเนียร์ทั้ง 4 สมการ สมการที่เป็นแบบคูณ เป็นความสัมพันธ์ที่ไม่เป็นความสัมพันธ์เชิงเส้นตรง Kennedy (1983) และ Steven (1996) ได้กล่าวไว้ว่า โมเดลแบบคูณนี้จะแปลงให้เป็นโมเดลเชิงเส้นตรงได้โดยการคำนวณหาค่าลอการิทึม โดยทั่วไปนิยมใช้ลอการิทึมธรรมชาติ (natural logarithm) ซึ่งอยู่ในรูปลอการิทึมฐาน e โดย e เป็นค่าคงที่มีค่าประมาณ 2.718 โมเดลที่ได้จากการคำนวณความถี่ที่คาดหวัง 4 แบบ สามารถทำให้เป็นโมเดลเชิงเส้นได้ดังนี้

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 $F_{ij} = \tau$ $\qquad$ ดังนั้น $\ln F_{ij} = \lambda$

จากสมการที่ 2 $F_{ij} = \tau\tau_i^a$ $\qquad$ $\ln F_{ij} = \lambda + \lambda_i^a$

จากสมการที่ 3 $F_{ij} = \tau\tau_i^a\tau_j^b$ $\qquad$ $\ln F_{ij} = \lambda + \lambda_i^a + \lambda_j^b$

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

โมเดลล็อกลิเนียร์ แบ่งออกเป็นหลายประเภท ตามลักษณะของการคำนวณค่าความถี่ที่คาดหวัง ในกรณีที่มีตัวแปร 2 ตัวแปร การคำนวณค่าความถี่ที่คาดหวังมี 4 ลักษณะ ได้สมการอธิบายค่าความถี่ที่คาดหวัง 4 รูปแบบนั้น เป็นที่มาของโมเดลล็อกลิเนียร์ทั่วๆ ไป (general log-linear model) โมเดลทั้ง 4 โมเดล คือโมเดลความน่าจะเป็นเท่า (mutual equiprobability model) โมเดลความน่าจะเป็นเท่ากันแบบมีเงื่อนไข (conditional equiprobability model) โมเดลอิสระต่อกัน (mutual independence model) และโมเดลอิ่มตัว (saturated model) (ทวีพร บุญวาณิช, 2541; Kennedy, 1983) ดังรายละเอียดต่อไปนี้

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

โมเดลนี้จะเป็นโมเดลที่ทำนายความถี่ที่คาดหวัง เป็นโมเดลที่แสดงว่า ความถี่ที่คาดหวังไม่ขึ้นกับอิทธิพลใดๆ ของตัวแปร หน่วยตัวอย่างมีคุณลักษณะที่เหมือนๆ กัน เป็นลักษณะที่คงที่ คือค่าเฉลี่ยรวมของความถี่ที่คาดหวังหรือ τ นั่นเอง การคำนวณความถี่ที่คาดหวังใช้จำนวนหน่วยตัวอย่างทั้งหมดเป็นฐานในการคำนวณจึงใช้สัญลักษณ์ [n] บอกถึงผลรวมความถี่ที่ใช้เป็นฐานในการคำนวณ ดังนั้นจะได้โมเดลล็อกลิเนียร์ที่มีลักษณะดังนี้

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

2) โมเดลความน่าจะเป็นเท่ากันแบบมีเงื่อนไข (Conditional Equiprobability Model)

โมเดลนี้ทำนายความถี่ที่คาดหวังเมื่อได้รับอิทธิพลหลักของ A ทำให้ความถี่ที่คาดหวังในแต่ละเซลล์ซึ่งควรจะเท่ากัน กลับมีค่าแตกต่างกัน ยกเว้นความถี่ที่คาดหวังทุกสดมภ์ในแต่ละแถวของ A มีค่าเท่ากัน และการคำนวณความถี่ที่คาดหวังใช้จำนวนหน่วยตัวอย่างในแถวของ A เป็นฐาน จึงใช้สัญลักษณ์ [A] แทนฐานข้อมูลที่ใช้ในการคำนวณความถี่ที่คาดหวัง จะได้โมเดลล็อก-ลิเนียร์ที่มีลักษณะดังนี้

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

3) โมเดลอิสระต่อกัน (Mutual Independence Model)

โมเดลนี้ทำนายความถี่ที่คาดหวังในกรณีที่ได้รับอิทธิพลหลักของ A อิทธิพลหลักของ B การคำนวณความถี่ที่คาดหวังใช้ผลรวมจำนวนตัวอย่างในแถวที่ i สดมภ์ที่ j เป็นฐาน จึงใช้

สัญลักษณ์ [AB] แทนฐานข้อมูลที่ใช้ในการคำนวณความถี่ที่คาดหวัง ดังนั้น จะได้โมเดลล็อก-ลิเนียร์ที่มีลักษณะดังนี้

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

4) โมเดลอิ่มตัว (saturated model)

โมเดลนี้ทำนายความถี่ที่คาดหวังที่ได้รับอิทธิพลหลักของ A อิทธิพลหลักของ B และปฏิสัมพันธ์ระหว่าง A กับ B ความถี่ที่คาดหวังเท่ากับความถี่ที่สังเกตได้เซลล์ ij นั่นคือ ความถี่ที่คาดหวังคำนวณโดยใช้จำนวนตัวอย่างในเซลล์ เป็นฐาน จึงใช้สัญลักษณ์ [AB] แทนฐานข้อมูลที่ใช้ในการคำนวณความถี่ที่คาดหวัง ดังนั้นจะได้โมเดลล็อกลิเนียร์ที่มีลักษณะดังนี้

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

3. **การประมาณค่าพารามิเตอร์** (Parameter Estimation)

การวิเคราะห์ล็อกลิเนียร์มี 4 โมเดล แต่ละโมเดลเทียบได้กับโมเดลการวิเคราะห์ความแปรปรวนดังที่กล่าวแล้ว พารามิเตอร์ คือ ค่าเฉลี่ยของความถี่ที่คาดหวังแต่ละเซลล์ ในที่นี้ คือ ค่ามัชฌิมเรขาคณิต (geometric mean) ของความถี่ที่คาดหวัง เมื่อใช้สัญญลักษณ์แทน ดังนี้

G.. แทน มัชฌิมเรขาคณิตทั้งหมดของความถี่ที่คาดหวัง (grand geometric mean)

$= (\Pi f_{ij}/ ij)^{1/ij}$

G_i^a แทน มัชฌิมเรขาคณิตของความถี่ที่คาดหวังของกลุ่มหรือในแต่ละระดับของ A

(group geometric mean) $= (\Pi f_{ij}/j)^{1/j}$

G_j^b แทน มัชฌิมเรขาคณิตของความถี่ที่คาดหวังของกลุ่มหรือในแต่ละระดับของ B

(group geometric mean) $= (\Pi f_{ij}/ j)^{1/i}$

โดยที่ Π หมายถึง ผลคูณ

นั่นคือ

<u>โมเดลล็อกลิเนียร์แบบคูณ</u>	<u>การประมาณค่าอิทธิพล</u>
$F_{ij} = \tau$	$\tau = G..$
$F_{ij} = \tau\tau_i^a$	$\tau = G..$, $\tau_i^a = G_i^a - G..$

$F_{ij} = \tau\tau_i^a\tau_j^b$ $\qquad$ $\tau = G..\ ,\ \tau_i^a = G_i^a - G..\ ,\ \tau_j^b = G_j^b - G..$

$F_{ij} = \tau\tau_i^a\tau_j^b\tau_{ij}^{ab}$ $\qquad$ $\tau = G..\ ,\ \tau_i^a = G_i^a - G..\ ,\ \tau_j^b = G_j^b - G..\ ,\ \tau_{ij}^{ab} = f_{ij} / \tau\tau_i^a\tau_j^b$

<u>โมเดลล็อกลิเนียร์แบบบวก</u> $\qquad$ <u>การประมาณค่าอิทธิพล</u>

$\ln F_{ij} = \lambda$ $\qquad$ $\lambda = \ln G..$

$\ln F_{ij} = \lambda + \lambda_i^a$ $\qquad$ $\lambda = \ln G..\ ,\ \lambda_i^a = \ln G_i^a - \ln G..$

$\ln F_{ij} = \lambda + \lambda_i^a + \lambda_j^b$ $\qquad$ $\lambda = \ln G..\ ,\ \lambda_i^a = \ln G_i^a - \ln G..,\ \lambda_j^b = \ln G_j^b - \ln G..$

$\ln F_{ij} = \lambda + \lambda_i^a + \lambda_j^b + \lambda_{ij}^{ab}$ $\qquad$ $\lambda = \ln G..\ ,\ \lambda_i^a = \ln G_i^a - \ln G..,\ \lambda_j^b = \ln G_j^b - \ln G..,$

$\lambda_{ij}^{ab} = \ln f_{ij} - \ln G_i^a - \ln G_j^b + \ln G..$

โดยมีข้อตกลงสำหรับการประมาณค่าเทอมอิทธิพลจากโมเดลแบบคูณ คือ

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

และมีข้อตกลงสำหรับการประมาณค่าเทอมอิทธิพลจากโมเดลแบบบวก คือ

$$\sum\lambda_i^a = 0\ ,\ \sum\lambda_j^b = 0\ ,\ \sum\sum\lambda_{ij}^{ab} = 0$$

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, 1990)

จากที่กล่าวมาจะเห็นว่าโมเดลล็อกลิเนียร์ กรณีตัวแปร 2 ตัวแปร มีโมเดลที่มีลักษณะต่างกัน 4 ลักษณะจึงจำเป็นที่จะต้องคัดเลือกโมเดลเพียงโมเดลเดียวที่มีความกลมกลืนกับข้อมูล และโมเดลที่เลือกเป็นโมเดลแบบประหยัด (parsimonious model) ดังหัวข้อต่อไป

4. การคัดเลือกโมเดลล็อกลิเนียร์ (Log-linear Models Selection)

การคัดเลือกโมเดลล็อกลิเนียร์ 4 โมเดลสำหรับการศึกษาตัวแปร 2 ตัวแปรนี้ให้เหลือเพียงโมเดลเดียวที่เหมาะสมกับการอธิบายความถี่ที่คาดหวังโดยการทดสอบความกลมกลืน (test of goodness of fit) ระหว่างความถี่ที่คาดหวัง (F_{ij}) กับข้อมูลเชิงประจักษ์หรือความถี่ที่สังเกตได้ (f_{ij}) จะได้โมเดลล็อกลิเนียร์ที่มีความสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์และง่ายที่สุดต่อการอธิบาย การทำนายความถี่ที่คาดหวังซึ่งสามารถพิจารณาความสัมพันธ์ หรือความแตกต่างของตัวแปรได้อย่างมีนัยสำคัญทางสถิติ เรียกว่าโมเดลประหยัด (parsimonious model) การคัดเลือกโมเดลล็อกลิเนียร์ที่เป็นโมเดลประหยัด มีขั้นตอนของการทดสอบความกลมกลืนของโมเดล 2 ขั้นตอนตามที่ Kennedy (1983) ได้อธิบายไว้คือ ขั้นตอนของการทดสอบความกลมกลืนระหว่างความถี่ที่คาดหวัง กับความถี่ที่สังเกตได้ และ ขั้นตอนการทดสอบนัยสำคัญของเทอมอิทธิพลที่เพิ่มเข้าไปในโมเดลล็อกลิเนียร์ที่ผ่านการทดสอบในขั้นตอนแรกทีละเทอมว่ามีอิทธิพลเพียงพอที่จะทำให้โมเดลล็อกลิเนียร์มีความสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ ขั้นตอนการทดสอบความกลมกลืนของโมเดล 2 ขั้นตอน ตามขั้นตอนของ Kennedy (ทวีพร บุญวาณิช, 2541; Kennedy,1983) อธิบายไว้ดังนี้

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

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

อัตราส่วนไลค์ลิฮูดไค-สแควร์ $$L^2 = 2\sum\sum(f_{ij}) \ln (f_{ij} / F_{ij})$$

ค่าสถิติทั้ง 2 ค่าอาจจะมีค่าต่างกันแต่ต่างกันไม่มากนัก โดยเฉพาะถ้าจำนวนกรณีหรือหน่วยตัวอย่างมีขนาดใหญ่ สถิติทั้ง 2 มีการแจกแจงแบบไค-สแควร์ และมีค่าใกล้เคียงกันมากทำให้ผลสรุปการทดสอบสมมุติฐานเหมือนกัน แต่สถิติอัตราส่วนไลค์ลิฮูดไค-สแควร์จะเป็นที่นิยมมากกว่า เพราะสถิติอัตราส่วนไลค์ลิฮูดไค-สแควร์สามารถแบ่งเป็นหลายค่า แต่ละค่าแสดงถึงอิทธิพลของแต่ละเทอมในโมเดล (ทวีพร บุญวาณิช, 2541; Goodman, 1971)

ขั้นตอนที่สอง ทดสอบนัยสำคัญของเทอมอิทธิพลที่เพิ่มเข้าไปในโมเดลล็อกลิเนียร์ที่ผ่านการทดสอบในขั้นตอนแรกทีละเทอมว่ามีอิทธิพลเพียงพอที่จะทำให้โมเดลล็อกลิเนียร์มีความกลมกลืนกับข้อมูลเชิงประจักษ์ โดยตั้งสมมุติฐานหลักว่า H_0 : อิทธิพลตัวใหม่ที่เพิ่มเข้าไปมีค่าเท่ากับศูนย์ การทดสอบทำได้ด้วยการคำนวณผลต่างของค่าสถิติ L^2 (ค่าผลต่างใช้สัญญลักษณ์ ΔL^2) ของโมเดลล็อกลิเนียร์ที่ทดสอบเป็นคู่ ๆ โดยโมเดลล็อกลิเนียร์แต่ละคู่ต้องมีลักษณะลดหลั่นเรียงกันไป คือ โมเดลล็อกลิเนียร์ที่มีความกลมกลืนกับข้อมูลแล้วนำมาทดสอบนัยสำคัญเทอมอิทธิพลเทอมสุดท้ายจะสามารถทดสอบคู่กับโมเดลอื่นที่มีเทอมอิทธิพลน้อยกว่าอยู่ 1 เทอม แต่โมเดลล็อกลิเนียร์ที่มีเทอมอิทธิพลน้อยกว่านั้น ต้องมีเทอมอิทธิพลเหมือนกับโมเดลล็อกลิเนียร์ที่ทดสอบนัยสำคัญด้วย เมื่อคำนวณค่า ΔL^2 นำมาเปรียบเทียบกับค่าไค-สแควร์จากตารางสถิติ คือ $\chi^2_{\alpha,df}$ ซึ่ง df คำนวณได้จากผลต่างขององศาอิสระของโมเดลล็อกลิเนียร์คู่ที่ทดสอบคู่นั้น ๆ แล้วพิจารณาเลือกผลการทดสอบที่ปฏิเสธสมมุติฐาน (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 6

ตารางที่ 6 ค่าสถิติอัตราส่วนไลค์ลิฮูดไค-สแควร์ และผลต่างของค่าสถิติอัตราส่วนไลค์ลิฮูดไค-สแควร์ของโมเดลล็อกลิเนียร์ กรณี 2 ตัวแปร

โมเดลที่	โมเดลแบบคูณ	H_o: Model กลมกลืนกับข้อมูล			H_o: 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

* α=.01

จากตารางที่ 6 โมเดลล็อกลิเนียร์ที่ทดสอบมี 4 โมเดล และการทดสอบในขั้นตอนแรกมีโมเดลที่คัดออก 2 โมเดล คือ โมเดลที่ 1 และ 2 ในขั้นตอนที่สองจึงพิจารณาเฉพาะโมเดลที่ 3 และ 4 ปรากฏว่าโมเดลที่ 3 เป็นโมเดลที่กลมกลืนกับข้อมูลที่ดีที่สุด เมื่อกำหนดให้ $\alpha = .01$

สำหรับการทดสอบโมเดลล็อกลิเนียร์ในขั้นตอนที่สองนั้น เป็นการทดสอบนัยสำคัญของเทอมอิทธิพลที่เพิ่มเข้ามา มีการทดสอบโมเดลที่ซับซ้อน (conditional test for nested model) ดังนี้ (Agresti,1984) คือ การเปรียบเทียบความแตกต่างของค่าสถิติ L^2 นั้นสามารถแสดงให้เห็นถึงการเปรียบเทียบโมเดลที่ซับซ้อน (nested model) ของโมเดลสองโมเดลได้ ตัวอย่างเช่นดังตารางที่ 6 โมเดลที่ 2 จะเป็นกรณีเฉพาะ (special case) ของโมเดลที่ 3 โดยที่เทอมอิทธิพลที่ต้องการศึกษาจะเป็นเทอมอิทธิพล τ^b ที่อยู่ในโมเดลที่ 3 เพิ่มขึ้นมา และมีสมมุติฐานว่า เทอมอิทธิพลที่เพิ่มเข้ามามีค่าเป็นศูนย์ ดังนั้นค่าสถิติที่เหมาะสมในการทดสอบครั้งนี้ก็คือ

$$L^2[(2)\,|\,(3)] = L^2[(2)] - L^2[(3)]$$
$$= 21.65 - 5.19 = 16.46 \quad ,df = 2\text{-}1 = 1$$

จะเห็นได้ว่าเทอมอิทธิพล τ^b ที่เพิ่มเข้ามานั้นมีนัยสำคัญทางสถิติ

หากมีตัวแปรจำนวนมากขึ้น การสร้างตารางการณ์จรจะมีขนาดใหญ่ และมีโมเดลล็อกลิเนียร์ที่มีเทอมอิทธิพลลดหลั่นกันจำนวนมากที่นำมาพิจารณาคัดเลือก เช่น มีตัวแปร 3

ตัวแปร จะได้โมเดลจำนวน 8 โมเดล ถ้ามีตัวแปร 4 ตัวแปร จะได้โมเดลจำนวน 16 โมเดล และถ้ามีตัวแปร 5 ตัวแปร จะได้โมเดลจำนวน 32 โมเดล ฉะนั้น ยิ่งมีตัวแปรมากการทดสอบความกลมกลืนยิ่งมีความซับซ้อนมากยิ่งขึ้น เพราะอาจมีโมเดลที่คาดว่าเหมาะสมจำนวนมากในการทดสอบในขั้นตอนแรก ทำให้เกิดความยากในการพิจารณาคัดเลือกโมเดล นักสถิติจึงพัฒนาวิธีการที่ทำให้การวิเคราะห์คัดเลือกโมเดลมีความยุ่งยากลดลง นั่นคือ Kennedy Ferthofer และ Lorimer (ทวีพร บุญวาณิช, 2541; Kennedy, 1983; Ferthofer and Lorimer, 1989) สรุปว่า ในกรณีมีตัวแปรหลายตัวแปร กระบวนการคัดเลือกโมเดลให้จัดกลุ่มโมเดลก่อน เช่น การวิเคราะห์ตัวแปร 3 ตัวแปรคือ 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 เรียกโมเดลที่ 2-4 ว่าโมเดล 1 ปัจจัย (1 factor) หรือโมเดลปัจจัยหลัก (main marginal effect model) โดยที่โมเดลที่ 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, BC เรียงกันไป 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 เรียกโมเดลล็อกลิเนียร์นี้ว่าโมเดล 3 ปัจจัย (3 factors) เนื่องจากมีโมเดลเดียวฉะนั้นโมเดลที่ 8 จึงเป็นโมเดลเต็มรูปปฏิสัมพันธ์อันดับสอง (full second order model) และเป็นโมเดลสุดท้ายของการวิเคราะห์ตัวแปร 3 ตัวแปร จะเห็นว่าโมเดลที่ 8 มีเทอมอิทธิพลครบทั้ง 8 เทอมจึงเรียกได้อีกชื่อว่า โมเดลอิ่มตัว (saturated model) ส่วนโมเดลที่ 1 - 7 มีเทอมอิทธิพลไม่ครบ เรียกโมเดลทั้ง 7 โมเดลนี้ว่าโมเดลไม่อิ่มตัว (unsaturated model) ดังตัวอย่างตามตารางที่ 7

ตารางที่ 7 การจัดกลุ่มโมเดลจาก 8 โมเดลให้เหลือเพียง 4 กลุ่ม

กลุ่มที่	ชื่อโมเดลล็อกลิเนียร์	เทอมอิทธิพล	โมเดลแบบคูณโมเดลที่
1	null model	ไม่มีเทอมอิทธิพลใดๆของตัวแปร	1:τ
2	1 factor	เทอมอิทธิพลหลัก A, B, C	2:$\tau\tau^{a}$
			3:$\tau\tau^{a}\tau^{b}$
			4:$\tau\tau^{a}\tau^{b}\tau^{c}$
3	2 factors	เทอมปฏิสัมพันธ์อันดับหนึ่ง	5:$\tau\tau^{a}\tau^{b}\tau^{c}\tau^{ac}$
			6:$\tau\tau^{a}\tau^{b}\tau^{c}\tau^{ac}\tau^{ab}$
			7:$\tau\tau^{a}\tau^{b}\tau^{c}\tau^{ac}\tau^{ab}\tau^{bc}$
4	3 factors	เทอมปฏิสัมพันธ์อันดับสอง	8:$\tau\tau^{a}\tau^{b}\tau^{c}\tau^{ac}\tau^{ab}\tau^{bc}\tau^{abc}$

ในการคัดเลือก 8 โมเดลนี้มีวิธีการลดจำนวนโมเดลในการทดสอบได้โดยการเลือกโมเดลเต็มรูปของแต่ละกลุ่ม คือ โมเดลที่ 1, 4, 7 และ 8 มาทำการทดสอบสมมุติฐานของความกลมกลืนของโมเดลตาม 2 ขั้นตอนที่กล่าวแล้ว เมื่อโมเดลใดเป็นโมเดลที่เหมาะสมจึงคัดเลือกโมเดลในกลุ่มเดียวกันอีกครั้งด้วยการทดสอบความกลมกลืนตามกระบวนคัดเลือก 2 ขั้นตอน วิธีการนี้จะทำให้ง่ายต่อการพิจารณามากขึ้น ซึ่งสามารถนำไปพัฒนาใช้กับการวิเคราะห์ตัวแปรมากกว่า 2 ตัวแปรได้ง่ายและสะดวกขึ้น

กระบวนการคัดเลือกโมเดล 2 ขั้นตอนนี้สามารถใช้กับการศึกษาทั้งแบบสมมาตรและแบบอสมมาตร แต่มีข้อแตกต่างกัน 3 ประการ คือ ประการแรกโมเดลในการพิจารณาสำหรับการศึกษาแบบสมมาตรเป็นโมเดลล็อกลิเนียร์ทุกโมเดลที่เป็นไปได้ที่ใช้ทำนายความถี่ที่คาดหวัง เรียกว่า โมเดลล็อกลิเนียร์ทั่ว ๆ ไป (general log-linear model) แต่โมเดลในการพิจารณาสำหรับการศึกษาแบบอสมมาตร นอกจากโมเดลล็อกลิเนียร์ทั่ว ๆ ไปแล้วยังมีโมเดลที่พัฒนาจากโมเดลล็อกลิเนียร์ทั่วๆไป เรียกว่า โมเดลโลจิท(logit model) โมเดลโลจิทแสดงความสัมพันธ์ระหว่างผลต่างของสัดส่วนของความถี่ในระดับของตัวแปรตามระดับต่างๆกับชุดอิทธิพลของตัวแปรอิสระ ประการที่สอง การศึกษาแบบอสมมาตรมีโมเดลสำหรับพิจารณาน้อยกว่าการศึกษาแบบสมมาตร และประการที่สาม การเลือกโมเดลในการศึกษาแบบอสมมาตรไม่เพียงแต่มีโมเดลที่เป็นโมเดลประหยัดและมีความกลมกลืนกับข้อมูลเชิง-

ประจักษ์ แต่โมเดลที่เลือกจำเป็นต้องเป็นโมเดลที่ระบุอิทธิพลของตัวแปรแตกต่างจากโมเดลโลจิทแบบศูนย์ (null logit model) ด้วย เช่น โมเดล 4 โมเดลของการศึกษา 2 ตัวแปรดังตารางที่ 6 หน้า 20 โมเดลที่ 1-4 เป็นโมเดลสำหรับการศึกษาแบบสมมาตร และโมเดลที่ 3 และ 4 เป็นโมเดลสำหรับการศึกษาแบบอสมมาตร เมื่อกำหนดให้ตัวแปร A เป็นตัวแปรอิสระ 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 เป็นโมเดลที่บอกความแตกต่างของสัดส่วนความถี่แต่ละระดับของตัวแปร B ในแต่ละระดับของตัวแปร A จึงเป็นโมเดลที่บอกอิทธิพลของ A ที่มีต่อ B

จากโมเดลจำนวนมากเมื่อผ่านกระบวนการคัดเลือกโมเดล กรณีการศึกษาแบบสมมาตร การคัดเลือกโมเดลควรพิจารณาคัดเลือกให้เหลือเพียงโมเดลเดียวที่กลมกลืนกับข้อมูลมากที่สุด แต่การศึกษาแบบอสมมาตรไม่จำเป็นต้องมีโมเดลเดียวที่มีความกลมกลืนกับข้อมูลเมื่อได้โมเดลที่คัดเลือกไว้แล้ว

5. การตรวจสอบความเหมาะสมของโมเดล (Examine of appropriated model)

เมื่อคัดเลือกโมเดลแล้ว ขั้นต่อไปคือการตรวจสอบโมเดล มีวิธีการตรวจสอบ 2 แบบ คือ ตรวจสอบด้วยการพิจารณาค่าเศษเหลือมาตรฐาน (standardize residaul, r_{ij}) และ ตรวจสอบด้วยการทดสอบนัยสำคัญของเทอมอิทธิพล ดังต่อไปนี้

5.1 การพิจารณาค่าเศษเหลือมาตรฐาน (standardize residual , r_{ij})

การคำนวณค่าเศษเหลือมาตรฐาน (r_{ij}) จากสูตร

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

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

5.2 การทดสอบนัยสำคัญของเทอมอิทธิพล

เป็นการทดสอบนัยสำคัญของเทอมอิทธิพลแต่ละเทอมในโมเดลที่คัดเลือกว่ามีมากเพียงพอหรือไม่ที่ทำให้โมเดลมีความกลมกลืนกับข้อมูล สถิติทดสอบ คือ z- test โดยมีสมมุติฐานหลัก คือ $H_0 : \lambda = 0$ และสูตรในการคำนวณ z มีดังนี้

$$z = \lambda / SE_{\lambda}$$

เนื่องจากโปรแกรมคอมพิวเตอร์ปัจจุบันมีพื้นฐานการคำนวณจากโมเดลแบบบวก เช่น การวิเคราะห์ความแปรปรวน และการวิเคราะห์การถดถอย ฉะนั้น โมเดลล็อกลิเนียร์ปกติอยู่ในรูปโมเดลแบบคูณสามารถปรับให้อยู่ในรูปโมเดลแบบบวกได้โดยใช้ลอการิทึม การทดสอบนัยสำคัญของอิทธิพลจึงนิยมใช้เทอมของ λ

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 โดยมีสมมุติฐานหลัก ดังนี้

การทดสอบนัยสำคัญของเทอมอิทธิพลหลัก A

H_0: $\lambda^{a}_{i} = 0$, i = 1, 2 , ..., a

สถิติที่ทดสอบได้แก่ z-test

$$z = \lambda^{a}_{i} / SE_{\lambda_i^a}$$

โดย $SE_{\lambda_i^a} = \dfrac{(\sum\sum 1/f_{ij})^{1/2}}{a}$

สำหรับการทดสอบนัยสำคัญของเทอมอิทธิพลปฏิสัมพันธ์ AB

H_0: $\lambda_{ij}^{ab} = 0$, i = 1 , 2 ,.....,a j = 1 , 2 , ...,b

สถิติที่ทดสอบ คือ

$$z = \lambda_{ij}^{ab} / SE_{\lambda_{ij}^{ab}}$$

โดย $SE_{\lambda_{ij}}{}^{ab} = \frac{(\sum\sum 1/f_{ij})^{1/2}}{ab}$

เกณฑ์ในการพิจารณา คือ ค่า z ที่คำนวณได้ต้องมากกว่า 1.96 (Norusis, 1994) เมื่อกำหนดระดับนัยสำคัญ 5 เปอร์เซ็นต์ หรือ เมื่อพิจารณาจากช่วงเชื่อมั่นในการประมาณค่าพารามิเตอร์ที่ระดับความเชื่อมั่น 95 เปอร์เซนต์ ค่าประมาณพารามิเตอร์ (λ) ต่ำสุดและสูงสุดหากไม่ ครอบคลุมค่าศูนย์ จะทำให้ผลการทดสอบปฏิเสธสมมุติฐานหลัก แสดงว่า พารามิเตอร์เทอมอิทธิพลมีค่าไม่เท่ากับศูนย์ หมายความว่าเทอมอิทธิพลนี้เพียงพอต่อการอธิบายหรือการทำนายความถี่ที่คาดหวัง ฉะนั้น เทอมอิทธิพลนี้จึงควรมีอยู่ในโมเดล

หากผู้วิจัยศึกษาความสัมพันธ์แบบสมมาตรที่เป็นการศึกษาความสัมพันธ์ระหว่างตัวแปรโดยมิได้กำหนดตัวแปรใดเป็นตัวแปรตามและตัวแปรอิสระ 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 แต่ถ้าผู้วิจัยสรุปได้ว่าตัวแปรไม่มีความสัมพันธ์กันก็ไม่จำเป็นต้องมีการศึกษาหาขนาด และทิศทางต่อไป หากผู้วิจัยศึกษาความสัมพันธ์แบบอสมมาตรที่กำหนดให้ตัวแปรหนึ่งเป็นตัวแปรตามและตัวแปรอื่น ๆ เป็นตัวแปรอิสระ และสรุปได้ว่าตัวแปรอิสระใดมีอิทธิพลต่อตัวแปรตาม ขั้นต่อไปเป็นการประมาณค่าพารามิเตอร์ลอการิทึมของอัตราส่วนแต้มต่อ คือ ค่าผลต่างของโอกาสในการเกิดเหตุการณ์สองเหตุการณ์ที่ใช้อธิบายความแตกต่างของระดับของตัวแปรอิสระที่ส่งผลต่อตัวแปรตาม และทำให้ทราบถึงขนาดอิทธิพลที่ส่งผลต่อตัวแปรตาม

6. การประมาณค่าพารามิเตอร์ลอการิทึมของอัตราส่วนแต้มต่อ (Log Odds Ratio)

การประมาณค่าลอการิทึมของอัตราส่วนแต้มต่อ ประมาณค่าได้จากโมเดลโลจิทที่พัฒนาจากโมเดลล็อกลิเนียร์ทั่วๆ ไป ที่ได้คัดเลือกว่าเป็นโมเดลที่มีความกลมกลืนกับข้อมูลอย่างเหมาะสม เมื่อประมาณค่าลอการิทึมอัตราส่วนแต้มต่อแล้ว ต้องเปลี่ยนค่าเป็นอัตราส่วนแต้มต่อเพื่อการแปลผลในขั้นตอนต่อไป

ค่าลอการิทึมของอัตราส่วนแต้มต่อ คือ ค่าที่แสดงผลต่างของสัดส่วนของความถี่ 2 ระดับของตัวแปรหนึ่ง ๆ ในที่นี้เป็นค่าที่บ่งบอกถึงค่าของอิทธิพลของตัวแปรหนึ่งที่มีต่ออีกตัวแปรหนึ่ง การประมาณค่าลอการิทึมของอัตราส่วนแต้มต่อประมาณค่าได้จากโมเดลโลจิทที่พัฒนาจากโมเดลล็อกลิเนียร์ทั่ว ๆ ไป ยกตัวอย่างเช่น การศึกษา 2 ตัวแปร คือ A และ B โดยที่ A มี 2 ค่า และ B มี 2 ค่า โมเดลล็อกลิเนียร์มี 4 โมเดล ตามตารางที่ 6 หน้า 21 คือ โมเดลที่ 1-4 และโมเดลสำหรับการศึกษาความสัมพันธ์แบบอสมมาตรมี 2 โมเดล คือ โมเดลที่ 3 และ 4 เมื่อกำหนดให้ B เป็นตัวแปรตาม และ A เป็นตัวแปรอิสระ หากโมเดล 4 มีความกลมกลืนกับข้อมูล แสดงว่า A มีอิทธิพลต่อ B และโมเดล 4 มีลักษณะดังนี้

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

สามารถนำโมเดลนี้มาพัฒนาเป็นโมเดลโลจิทเพื่อประมาณค่าลอการิทึมของอัตราส่วนแต้มต่อของตัวแปร A ตามที่ Goodman (1972) เสนอไว้และอธิบายโดย Knoke และ Burke (1980) สรุปได้ว่า โมเดลโลจิทมีเกณฑ์ที่ใช้ในการพิจารณา คือ สัดส่วนของความถี่ของตัวแปรตาม ซึ่งต่างจากโมเดลล็อกลิเนียร์ทั่ว ๆ ไป ที่ใช้ความถี่เป็นเกณฑ์ในการพิจารณา ยกตัวอย่างเช่น กรณีโมเดลตามสมการที่ (1) พัฒนาเป็นโมเดลโลจิท ดังนี้

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

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

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

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

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

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

$$= 2\lambda_1^{b} + 2\lambda_{i1}^{ab}$$

$$\Phi_i^{b} = \beta_1^{b} + \beta_{i1}^{ab} \quad \text{....................(2)}$$

เมื่อ $\Phi_i^{b} = \ln(F_{i1} / F_{i2})$

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

$\beta_{i1}^{ab} = 2\lambda_{i1}^{ab}$

สมการ (2) เรียกว่าโมเดลโลจิท โดย $-\infty \leq \beta_1^{b} \leq +\infty$, $-\infty \leq \beta_{ij}^{ab} \leq +\infty$ 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_i^{b} มีค่าเพิ่มมากขึ้น และหากมีค่าลบ แสดงว่า อิทธิพลของเทอมนั้น ๆ มีผลทำให้การทำนายค่า Φ_i^{b} มีค่าน้อยลง (ทวีพร บุญวาณิช, 2541; Knoke และ Burke, 1980; Norusis, 1992)

เมื่อประมาณค่าลอการิทึมอัตราส่วนแต้มต่อแล้วต้องเปลี่ยนค่าเป็นอัตราส่วนแต้มต่อเพื่อการแปลผลต่อไป เนื่องจากค่าอัตราส่วนแต้มต่อ คือ สัดส่วนของโอกาสการเกิดเหตุการณ์สองเหตุการณ์ ฉะนั้น อัตราส่วนแต้มต่อมีค่ามากกว่าศูนย์เสมอ ถ้าอัตราส่วนแต้มต่อมีค่าเท่ากับ 1 แสดงว่า โอกาสของการเกิดเหตุการณ์ 2 เหตุการณ์เท่า ๆ กัน ถ้าอัตราส่วนแต้มต่อมีค่ามากกว่า 1 แสดงว่า โอกาสการเกิดเหตุการณ์หนึ่งมีมากกว่าเหตุการณ์อีกเหตุการณ์หนึ่ง และถ้าอัตราส่วนแต้มต่อมีค่าน้อยกว่า 1 แสดงว่า โอกาสการเกิดเหตุการณ์หนึ่งมีน้อยกว่าเหตุการณ์อีกเหตุการณ์หนึ่ง (Knoke และ Burke, 1980; Norusis, 1992)

ตอนที่ 2 การวิเคราะห์ล็อกลิเนียร์มาตรอันดับ

การวิเคราะห์ล็อกลิเนียร์มาตรอันดับ สามารถวิเคราะห์ได้ในกรณีที่ตัวแปรตั้งแต่หนึ่งตัวขึ้นไปเป็นตัวแปรมาตรอันดับ ดังนั้นจึงอาจมีการนำตัวแปรมาตรอันดับเข้ามาศึกษากับตัวแปรมาตรนามบัญญัติ หรือ เป็นการศึกษาในตัวแปรที่เป็นมาตรอันดับทั้งหมดก็ได้ ในที่นี้จะขอนำเสนอการวิเคราะห์ล็อกลิเนียร์ในกรณีที่เป็นตัวแปรมาตรอันดับทั้งหมด ซึ่งการวิเคราะห์ล็อกลิเนียร์มาตรอันดับนั้นมีขั้นตอนการวิเคราะห์ที่เหมือนกับการวิเคราะห์ล็อกลิเนียร์มาตรนามบัญญัติ เพียงแต่มีการศึกษาถึงโมเดลเกี่ยวเนื่องเพิ่มขึ้น ดังจะได้กล่าวถึงต่อไป

ในการวิเคราะห์ล็อกลิเนียร์มาตรอันดับ ในตารางการณ์จรขนาด 2 มิตินั้นจะได้โมเดลล็อกลิเนียร์ที่เป็นโมเดลอิสระ (independence model) (Agresti, 1984) คือ

$$\ln F_{ij} = \lambda + \lambda_i^{X} + \lambda_j^{Y}$$

เมื่อ F_{ij} = ความถี่ที่คาดหวัง

λ = ลอการิทึมของค่าเฉลี่ยรวมของความถี่ที่คาดหวัง

λ_i^X = ลอการิทึมของอิทธิพลหลักจากตัวแปร X

λ_j^Y = ลอการิทึมของอิทธิพลหลักจากตัวแปร Y

การกำหนดตัวแปรในโมเดลนี้ ตัวแปรทั้งด้านแถว (rows) และด้านสดมภ์ (columns) จะเป็นตัวแปรมาตรอันดับ โดยจะมีเซตของ u_i $\{u_i\}$ และเซตของ v_j $\{v_j\}$ ถูกกำหนดให้เป็นคะแนนของระดับการแปรค่าของตัวแปรทางด้านแถวของตัวแปร X และคะแนนของระดับการแปรค่าของตัวแปรทางด้านสดมภ์ของตัวแปร Y แต่ละด้าน ตามลำดับ ในการกำหนดคะแนนในแถวแต่ละแถว และสดมภ์แต่ละสดมภ์นั้น จะกำหนดให้คะแนนของแถวระดับที่ 1 น้อยกว่าคะแนนในแถวระดับที่ 2 น้อยกว่าคะแนนในแถวระดับที่ 3 ไปจนกระทั่งน้อยกว่าคะแนนในแถวระดับที่ r นั่นคือ $u_1 < u_2 < u_3 <\ldots< u_r$ และคะแนนในสดมภ์ระดับที่ 1 น้อยกว่าคะแนนในสดมภ์ระดับที่ 2 น้อยกว่าคะแนนในสดมภ์ระดับที่ 3 ไปจนกระทั่งน้อยกว่าคะแนนในสดมภ์ที่ c นั่นคือ $v_1 < v_2 < v_3 <\ldots< v_c$ การกำหนดคะแนนให้แต่ละระดับการแปรค่าของตัวแปรด้านแถวและสดมภ์อาจจะใช้จุดกึ่งกลาง (midpoints) ของค่าแต่ละระดับ เช่น ทางด้านแถวมีการแปรค่าระดับคะแนนของตัวแปรเป็น 5 ระดับ จะมีคะแนน $u_1 < u_2 < u_3 <\ldots< u_r$ มีการกำหนดคะแนน u_i ให้เท่ากับจุดกึ่งกลางของแต่ละระดับดังตารางที่ 8 หรืออาจจะกำหนดคะแนนในแต่ละระดับของด้านแถวด้วยลำดับที่แต่ละแถว เช่น แถวที่ 1 จะมี $u_i = 1$ แถวที่ 2 จะมี $u_i = 2$ ไปเรื่อยๆ ถ้ามี 5 แถว คะแนนจะเป็น 1,2,3,4,5 หรือ อาจจะกำหนดให้เป็น 10,20,30,40,50 แทนคะแนนแต่ละแถวได้เช่นเดียวกัน (Agresti, 1984; 1996)

ตารางที่ 8 คะแนน u_i ของแต่ละแถว

ระดับการแปรค่าของตัวแปรทางด้านแถว	คะแนน u_i
0	0
<1	0.5
1-2	1.5
3-5	4.0
≥6	7.0

จากการวิเคราะห์ล็อกลิเนียร์สำหรับโมเดลอิสระจะเป็นในกรณีที่มีอิทธิพลหลักจากตัวแปรอิสระเท่านั้น แต่ในการวิเคราะห์ที่มีการศึกษาถึงอิทธิพลปฏิสัมพันธ์ระหว่างตัวแปร หรือมีการศึกษาถึงโมเดลเกี่ยวเนื่อง (uniform association model) ในตัวแปรมาตรอันดับ จะได้โมเดลคือ

$$\ln F_{ij}=\lambda+\lambda_i^X+\lambda_j^Y+\beta(u_i-\bar{u})(v_j-\bar{v})$$

เมื่อ $\sum\lambda_i^X=\sum\lambda_j^Y=0$ ในโมเดลนี้ จะมีค่าพารามิเตอร์เพิ่มมาจากโมเดลอิสระหนึ่งค่า คือ ค่า β ซึ่งเป็นเทอมเกี่ยวเนื่อง (assosiation term) ดังนั้นจะได้ค่าองศาอิสระ (degrees of freedom) เท่ากับ (r-1)(c-1)-1 = rc-r-c สำหรับทดสอบความกลมกลืนของโมเดล (goodness of fit) ซึ่งจะเหมือนกับ โมเดลอิ่มตัว (saturated model) ของการวิเคราะห์ล็อกลิเนียร์สำหรับตัวแปรมาตรนามบัญญัติ คือ $\ln F_{ij}=\lambda+\lambda_i^X+\lambda_j^Y+\lambda_{ij}^{XY}$ ซึ่งเทอมของ λ_{ij}^{XY} มีโครงสร้างมาจาก $\beta(u_i-\bar{u})(v_j-\bar{v})$

ค่า β จะอธิบายถึงความสัมพันธ์ระหว่าง X และ Y โมเดลอิสระจะมีค่า β = 0 เทอมของ $\beta(u_i-\bar{u})(v_j-\bar{v})$ สะท้อนให้เห็นถึงการเบี่ยงเบนจาก $\ln F_{ij}$ จากโมเดลอิสระ ถ้าค่า β มีค่ามากกว่า 0 จะคาดได้ว่าค่า X มีขนาดใหญ่ และ Y มีขนาดใหญ่ หรือ อีกกรณีหนึ่งคือ ค่า X มีขนาดเล็ก และค่า Y มีขนาดเล็ก (ซึ่งในกรณีของโมเดลอิสระจะไม่สามารถคาดได้) ในทางกลับกัน ถ้า β มีค่าน้อยกว่า 0 จะแสดงให้เห็นว่าการเบี่ยงเบนจาก $\ln F_{ij}$ จากโมเดลอิสระ คือ ค่า X มีขนาดใหญ่ และ Y มีขนาดเล็ก หรือในอีกกรณีหนึ่งคือ ค่า X มีขนาดเล็ก และ Y มีขนาดใหญ่ โดยที่ค่า $(u_i-\bar{u})(v_j-\bar{v})$ นั้นมีค่าตรงกับ -1, 0 หรือ 1

ในกรณีที่แถว a มีค่าน้อยกว่าแถวที่ b และ สดมภ์ที่ c มีค่าน้อยกว่าสดมภ์ที่ d (ในแต่ละคู่) จะได้ค่า $\beta(u_i-\bar{u})(v_j-\bar{v})$ ว่า

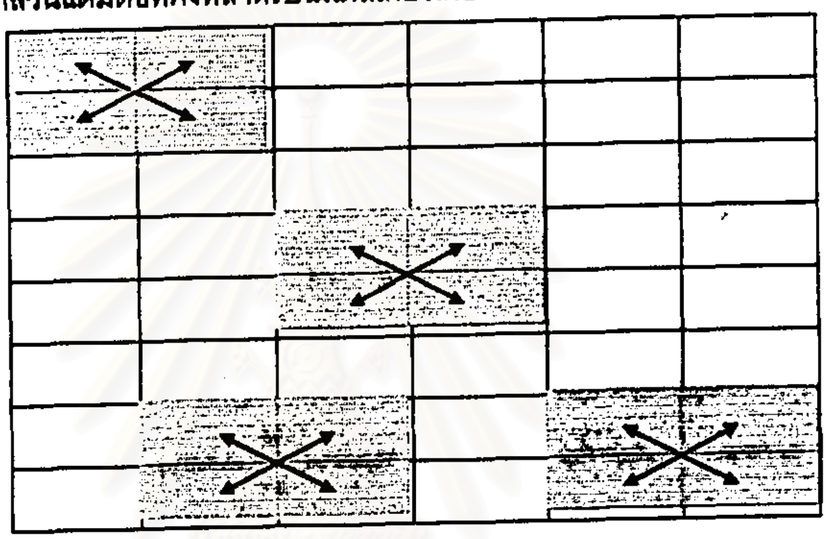
$$\ln\left(\frac{F_{ac}F_{bd}}{F_{ad}F_{bc}}\right)=\beta(u_i-\bar{u})(v_j-\bar{v})$$

Goodman และ Haberman (Agresti, 1984) ได้กำหนดโมเดลล็อกลิเนียร์สำหรับตัวแปรมาตรอันดับในเทอมของ (r-1)(c-1) ในการหาค่าอัตราส่วนแต้มต่อ (local odds ratios) ว่า

$$\theta_{ij}=\frac{\left(F_{ij}F_{i+1,j+1}\right)}{\left(F_{i,j+1}F_{i+1,j}\right)}$$

ในกรณีที่ใช้สำหรับแถวที่ติดกัน และสดมภ์ที่ติดกัน เมื่อ $1 \le i \le r-1$ และ $1 \le j \le c-1$ Goodman กล่าวว่า โมเดลเกี่ยวเนื่องนี้จะได้ค่า θ_{ij} ทุกตัวที่มีค่าเท่ากัน นั่นคือ ควรให้คะแนน u_i และ v_j มีคะแนนเหมือนกับเป็นอันตรภาค และ $u_2 - u_1 = u_3 - u_2 = ... = u_r - u_{r-1}$ และ $v_2 - v_1 = v_3 - v_2 = ... = v_c - v_{c-1}$ เป็นผลให้ θ_{ij} ทุกตัวมีค่าเท่ากับ exp(β) และค่า $\ln\theta_{ij}$ ทั้งหมดมีค่าเท่ากับ β จึงจะสามารถรายงานผลค่าลอการิทึมของอัตราส่วนแต้มต่อ (local log odds ratio)ได้ ดังภาพที่ 2

ภาพที่ 2 ค่าอัตราส่วนแต้มต่อที่คงที่สำหรับโมเดลเกี่ยวเนื่อง



ตัวอย่างที่แสดงถึงการคำนวณค่า β ดังเช่นข้อมูลจากตารางที่ 9 เมื่อได้ความถี่ที่สังเกตได้มาแล้วจะทำการคำนวณค่าความถี่ที่คาดหวัง ในการคำนวณความถี่ที่คาดหวังสำหรับในโมเดลอิสระจะหาได้จากผลคูณของผลรวมความถี่ที่สังเกตได้ในแต่ละแถว กับผลรวมของความถี่ที่สังเกตได้ในแต่ละสดมภ์หารด้วยผลรวมของความถี่ทั้งหมด (หรือจำนวนกลุ่มตัวอย่าง) ดังสูตร $(n_{i.})(n_{.j})/n$ ส่วนการคำนวณค่าความถี่ที่คาดหวังของโมเดลเกี่ยวเนื่อง (uniform association model) นั้น จะได้มาจากการคำนวณด้วยโปรแกรมคอมพิวเตอร์ โดยการคำนวณทวนซ้ำ ตามสูตรดังต่อไปนี้ ซึ่งจะได้ค่าความถี่ที่คาดหวังตามตารางที่ 9

$$F_{ij}^{(t+1)} = \left(\frac{n_{i.}}{F_{i.}}\right) F_{ij}^{(t)}$$

$$F_{ij}^{(t+2)} = \left(\frac{n_{.j}}{F_{.j}^{(t+1)}}\right) F_{ij}^{(t+1)}$$

$$F_{ij}^{(t+3)} = F_{ij}^{(t+2)} \left(\frac{\sum\sum u_a^* v_b^* n_{ab}}{\sum\sum u_a^* u_b^* F_{ab}^{(t+2)}}\right)^{u_i^* v_j^*} \left(\frac{\sum\sum (1-u_a^* v_b^*) n_{ab}}{\sum\sum (1-u_a^* v_b^*) F_{ab}^{(t+2)}}\right)^{1-u_i^* v_j^*}$$

เมื่อ $\{u_i^*\}$ และ $\{v_j^*\}$ เป็นคะแนนที่ $0 \le u_i^* \le 1$ และ $0 \le v_j^* \le 1$

จากตารางที่ 9 จะได้ค่า θ_{ij} = exp(β) = 1.18 จากค่าความถี่ที่คาดหวังในโมเดลเกี่ยวเนื่อง

$$1.18 = \frac{(62.5 \times 30.9)}{(26.2 \times 62.9)} = \ldots\ldots = \frac{(35.3 \times 16.7)}{(13.7 \times 36.6)}$$

การประมาณค่าพารามิเตอร์ (β) ทำได้โดย หาค่าลอการิทึมของ θ_{ij} ดังสูตร $\ln \theta_{ij} = \beta$ ดังนั้น β มีค่าเท่ากับ 0.163

ตารางที่ 9 ค่าความถี่ที่คาดหวังในโมเดลอิสระ และในโมเดลเกี่ยวเนื่อง

X	Y 1	Y 2	Y 3	รวม
1	61	28	7	96
	(55.3)[a]	(29.7)	(11.0)	
	(62.5)[b]	(26.2)	(7.3)	
2	68	23	13	104
	(59.9)	(32.2)	(12.0)	
	(62.9)	(30.9)	(10.2)	
3	58	40	12	110
	(63.3)	(34.0)	(12.7)	
	(61.0)	(35.3)	(13.7)	
4	53	38	16	107
	(61.6)	(33.1)	(12.3)	
	(53.7)	(36.6)	(16.7)	

[a] สำหรับค่าความถี่ที่คาดหวัง (F_{ij}) ในโมเดลอิสระ

[b] สำหรับค่าความถี่ที่คาดหวัง (F_{ij}) ในโมเดลเกี่ยวเนื่อง

การทดสอบความกลมกลืนของโมเดล

การทดสอบความกลมกลืนของโมเดลสามารถใช้การคำนวณค่า χ^2 ของ Pearson หรือใช้การทดสอบความกลมกลืนโดยใช้อัตราส่วนไลค์ลิฮูดไค-สแควร์ (likelihood ratio chi-square) ดังสูตรต่อไปนี้

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

อัตราส่วนไลค์ลิฮูดไค-สแควร์ $$L^2 = 2\sum\sum (f_{ij}) \ln (f_{ij} / F_{ij})$$

ในการคำนวณค่าทั้งสองค่านี้ จะมีค่าไม่แตกต่างกันมากนัก ถ้ากลุ่มตัวอย่างมีขนาดใหญ่ $(n \to \infty)$

จากตารางที่ 9 จะได้ค่าอัตราส่วนไลค์ลิฮูดไค-สแควร์ของโมเดลอิสระแทนด้วย L^2(I) มีค่าเท่ากับ 10.88 โดยมีค่าองศาอิสระเท่ากับ 6 และ L^2 ของโมเดลเกี่ยวเนื่องแทนด้วย L^2(U) มีค่าเท่ากับ 4.59 โดยมีค่าองศาอิสระเท่ากับ 5 ดังตารางที่ 10

ตารางที่ 10 ค่าความกลมกลืนของโมเดลอิสระ โมเดลเกี่ยวเนื่อง และโมเดลอิสระที่ให้มีความเกี่ยวเนื่อง

โมเดล	L^2	df
Independence	10.88	6
Uniform association	4.59	5
Indepence, given uniform association	6.29	1

การทดสอบเงื่อนไขของความเป็นอิสระ

การทดสอบความเป็นอิสระของโมเดลจะทดสอบโดยกำหนดสมมติฐานศูนย์ว่า ค่า β เท่ากับ 0 (H_0: β=0) จะสรุปในค่าของ L^2 ว่า

$L^2(I|U) = L^2(I) - L^2(U)$; df = 1

ในตารางที่ 2 นี้ จะได้ L^2(I U) =10.88-4.59 = 6.29 ค่า L^2 ที่ได้คือ ค่า L^2 ที่ได้รับเมื่อเพิ่มพารามิเตอร์ของความเกี่ยวเนื่อง (association parameter (β)) เข้ามาในโมเดลอิสระ ทำการ

ทดสอบที่ค่านัยสำคัญ P = 0.01 สำหรับการทดสอบ H_0: $\beta=0$ ดังนั้นจะเห็นได้ว่าความสัมพันธ์ของเหตุการณ์ที่ได้จะดีกว่าที่ได้รับจากโมเดลอิสระแต่เพียงอย่างเดียว

การคำนวณค่าอัตราส่วนแต้มต่อ

ในกรณีที่ทดสอบแล้วว่าโมเดลมีความเกี่ยวเนื่องกัน จึงสรุปได้ว่าตัวแปรมีความสัมพันธ์กัน แล้วจึงมาคำนวณค่าอัตราส่วนแต้มต่อ การคำนวณอัตราส่วนแต้มต่อแต่ละคู่นั้น เช่น แถว a และ แถว b โดยกำหนดให้ a<b และในสดมภ์ c และ d กำหนดให้ สดมภ์ c<d จะได้ค่าลอการิทึมของอัตราส่วนแต้มต่อดังนี้

$$
\begin{aligned}
\ln\frac{F_{ac}F_{bc}}{F_{ad}F_{bc}} &= \left[\lambda+\lambda_a+\lambda_c+\beta(u_a-\bar{u})(v_c-\bar{v})+\lambda+\lambda_b+\lambda_d+\beta(u_b-\bar{u})(v_d-\bar{v})\right]- \\
&\quad \left[\lambda+\lambda_a+\lambda_d+\beta(u_a-\bar{u})(v_d-\bar{v})+\lambda+\lambda_b+\lambda_c+\beta(u_b-\bar{u})(v_c-\bar{v})\right] \\
&= \left[\beta(u_a-\bar{u})(v_c-\bar{v})+\beta(u_b-\bar{u})(v_d-\bar{v})\right]- \\
&\quad \left[\beta(u_a-\bar{u})(v_d-\bar{v})+\beta(u_b-\bar{u})(v_c-\bar{v})\right] \\
&= \beta(u_a v_c+u_b v_d)-\beta(u_a v_d+u_b v_c) \\
&= \beta(u_b-u_a)(v_d-v_c)
\end{aligned}
$$

จะได้ว่า $\ln\frac{F_{ij}F_{i+1,j+1}}{F_{i,j+1}F_{i+1,j}} = \beta(u_{i+1})(v_{j+1}-v_j)$

$$= \beta$$

โมเดลโลจิทสำหรับการวิเคราะห์ล็อกลิเนียร์มาตรอันดับ

ในการพิจารณาคัดเลือกโมเดลของการวิเคราะห์แบบอสมมาตรนั้น จะมีโมเดลล็อกลิเนียร์ทั่วๆ ไป และมีโมเดลที่พัฒนาจากโมเดลล็อกลิเนียร์ทั่วๆ ไป เรียกว่า โมเดลโลจิท (logit model) โมเดลโลจิทแสดงความสัมพันธ์ระหว่างผลต่างของสัดส่วนของความถี่ในระดับของตัวแปรตามสองระดับ หรือ ตัวแปรตามหลายระดับกับชุดอิทธิพลของตัวแปรอิสระ (ทวีพร บุญวาณิช, 2541) ในกรณีของการวิเคราะห์ในครั้งนี้เป็นการวิเคราะห์ล็อกลิเนียร์มาตรอันดับ จึงมีการใช้โมเดลโลจิทสำหรับการวิเคราะห์ล็อกลิเนียร์มาตรอันดับเกิดขึ้น ซึ่งจะกล่าวถึงรายละเอียดของโมเดลโลจิทสำหรับการวิเคราะห์ล็อกลิเนียร์มาตรอันดับโดยสังเขป ดังนี้

กรณีที่ X และ Y เป็นตัวแปรมาตรอันดับ โดยกำหนดให้ X เป็นตัวแปรอิสระ และ Y เป็นตัวแปรตาม และในกรณีที่ตัวแปรตามมีการแปรค่าได้ C ระดับ จะได้โมเดลโลจิทจำนวน C-1 สมการ เมื่อให้ $L_{j(i)}$ เป็นโมเดลโลจิทที่มีการแปลค่าด้านแถวและด้านสดมภ์ เป็น i และ j ตามลำดับ จะได้โมเดลโลจิท (Agresti, 1984) ดังนี้

$$L_{j(i)} = \log\left[\frac{\left(F_{i,j+1} + \dots + F_{ic}\right)}{\left(F_{i1} + \dots + F_{ij}\right)}\right]$$

ดังนั้นโมเดลโลจิทจะประกอบไปด้วย $\{L_{1(i)}, L_{2(i)}, \dots, L_{c-1(i)}\}$ โดยที่คะแนนของ $\{u_i\}$ แต่ละค่าจะกำหนดอยู่ในด้านแถว

ในกรณีที่มีการศึกษาถึงโมเดลเกี่ยวเนื่อง (uniform association model) เกิดขึ้นมานั้น จะสามารถสร้างเป็นโมเดลโลจิทได้ (Agresti, 1984) ดังนี้

$$L_{j(i)} = \alpha_j + \beta_j(u_i - \bar{u}) \quad \text{โดยที่} \quad i = 1,\dots,r$$

โมเดลข้างต้นนี้คือ โมเดลการถดถอยแบบโลจิทในกรณีตัวแปรตามแบ่งเป็น 2 ระดับ ถ้าตัวแปรตามแบ่งเป็น 3 ระดับ จะได้โมเดลโลจิทจำนวน 3-1 = 2 โมเดล กำหนดให้โมเดล 2 โมเดลนี้มีค่า j = 1 และ j = 2 ตามลำดับ พารามิเตอร์ที่ประมาณค่าออกมาได้คือ β_1 และ β_2 ดังนั้นสามารถเขียนเป็นโมเดลโลจิทได้ดังนี้

$$L_{1(i)} = \alpha_1 + \beta_1(u_i - \bar{u}) \quad \text{และ} \quad L_{2(i)} = \alpha_2 + \beta_2(u_i - \bar{u})$$

จากโมเดลทั้งสองโมเดล สามารถเขียนในรูปโมเดลหนึ่งโมเดล ได้ดังนี้

$$L_{j(i)} = \alpha_j + \beta_j(u_i - \bar{u}) \quad \text{โดยที่ } 1 \le i \le r,\ 1 \le j \le c-1 \text{ เมื่อ } \alpha_j \text{ เป็นค่าคงที่}$$

ส่วนการประมาณค่าพารามิเตอร์ที่ประมาณค่าเป็นค่า β นั้น สามารถรายงานผลในรูปของค่าลอการิทึมอัตราส่วนแต้มต่อ และในด้านการแปลผลการประมาณค่าพารามิเตอร์ β นั้น ถ้าค่า β มีค่าเท่ากับศูนย์แล้ว สามารถบอกได้ว่า X และ Y เป็นอิสระต่อกัน ถ้าค่า β มีค่ามากกว่า

ศูนย์ แสดงว่าถ้าตัวแปรอิสระมีจำนวนที่เพิ่มมากขึ้น จะเป็นผลทำให้ตัวแปรตามมีจำนวนที่เพิ่มมากขึ้นด้วยเช่นกัน แต่ถ้าในกรณีที่ค่า β มีค่าน้อยกว่าศูนย์ แสดงว่า ถ้าตัวแปรอิสระมีจำนวนที่เพิ่มมากขึ้น จะเป็นผลทำให้ตัวแปรตามมีจำนวนที่ลดน้อยลง และในทางกลับกัน ถ้าตัวแปรอิสระมีจำนวนที่ลดน้อยลง จะเป็นผลทำให้ตัวแปรตามมีจำนวนที่เพิ่มมากขึ้น

ส่วนการแปลความหมายของสัมประสิทธิ์ของโมเดลโลจิทต้องแปลความหมายในรูปของอัตราส่วนของแต้มต่อ (odds ratio) ซึ่งเป็นการเปลี่ยนแปลงของแต้มต่อ (odds) ตามการเปลี่ยนแปลงของตัวแปร โดยใช้ค่า e ยกกำลังด้วย β ซึ่งเป็นค่าแต้มต่อที่เปลี่ยนแปลง เมื่อตัวแปรอิสระตัวที่ i มีค่าเปลี่ยนไป 1 หน่วย ถ้า β มีค่าเป็นบวก เทอมนี้จะมีค่ามากกว่า 1 หมายความว่าค่าของแต้มต่อเพิ่มขึ้น ถ้า β มีค่าเป็นลบ เทอมนี้จะมีค่าน้อยกว่า 1 หมายความว่าค่าของแต้มต่อจะลดลง และถ้า β มีค่าเป็นศูนย์ เทอมนี้จะมีค่าเท่ากับ 1 หมายความว่าค่าของแต้มต่อไม่เปลี่ยนแปลง ดังนั้นในการแปลความหมายของสัมประสิทธิ์ของโมเดลโลจิทจึงเป็นการแปลความหมายของ ขนาดการเปลี่ยนแปลงของตัวแปรตามในรูปของอัตราส่วนแต้มต่อ เมื่อตัวแปรอิสระเปลี่ยนไป 1 หน่วย สำหรับการแปลค่าแต้มต่อในกรณีที่ตัวแปรตามแปรค่าได้ 3 ค่า (1, 2 และ 3) และเป็นตัวแปรมาตรอันดับนั้น จะพิจารณาจากโมเดลโลจิททั้งสอง (C - 1) กล่าวคือ ในโมเดลโลจิทที่ 1 จะเป็นการเปรียบเทียบแต้มต่อระหว่างค่าของตัวแปรระดับที่ 1 กับกลุ่มของค่าตัวแปรระดับที่ 2 และ 3 รวมกัน ส่วนโมเดลโลจิทที่ 2 จะเป็นการเปรียบเทียบแต้มต่อระหว่างกลุ่มของตัวแปรระดับที่ 1 และ 2 รวมกันกับค่าของตัวแปรระดับที่ 3 (Agresti, 1984; DeMaris, 1992; Hosmer and Lemmeshow, 1989)

ตอนที่ 3 งานวิจัยที่เกี่ยวข้องกับการศึกษาวิจัยเรื่องตัวแปรที่สัมพันธ์กับการศึกษาต่อและไม่ศึกษาต่อของนักเรียน

จากการศึกษางานวิจัย 14 เรื่อง พบว่า ตัวแปรที่สัมพันธ์กับการศึกษาต่อและไม่ศึกษาต่อมีตัวแปรที่สำคัญอยู่ 12 ตัวแปรด้วยกัน ได้แก่ ตัวแปรความต้องการศึกษาต่อของนักเรียน ความมุ่งหวังของผู้ปกครองที่จะให้บุตรได้รับการศึกษาขั้นสูง ผลสัมฤทธิ์ทางการเรียนของนักเรียน อาชีพของบิดามารดาหรือผู้ปกครอง รายได้ของบิดามารดาหรือผู้ปกครอง ระดับการศึกษาของบิดามารดาหรือ ผู้ปกครอง สถานภาพสมรสของบิดามารดา ภูมิลำเนาของบิดามารดา จำนวนบุตรในครอบครัว จำนวนบุตรที่ได้รับการศึกษาขั้นสูงหรือการมีพี่ศึกษาอยู่ในระดับขั้นมัธยมศึกษาหรือสูงกว่าบิดา ระยะทางจากบ้านถึงโรงเรียน และความสะดวกในการคมนาคม ดังตารางที่ 11

ตารางที่ 11 ตัวแปรที่สัมพันธ์กับการศึกษาต่อและไม่ศึกษาต่อของนักเรียน

ผู้วิจัย / ตัวแปร	จรูญ	ชดทิพย์	ดำรง	เทียมจันทร์	ธวัช	บุญเพ็ง	ประเสริฐ	ภัทรธิรา	วิศิษฐ์	สวลี	อุดม	Blake	William	Worthington
ความต้องการศึกษาต่อของนักเรียน				/	/			/						
ความมุ่งหวังของผู้ปกครองที่จะให้บุตรได้รับการศึกษาขั้นสูง			/	/	/									
ผลสัมฤทธิ์ทางการเรียนของนักเรียน	/		/	/	/	/				/				/
อาชีพของบิดามารดาหรือผู้ปกครอง	/	/	/	/				/	/	/	/	/	/	/
รายได้ของบิดามารดาหรือผู้ปกครอง		/	/	/			/		/	/		/		/
การศึกษาของบิดามารดาหรือผู้ปกครอง	/	/		/		/			/	/	/	/		/
สถานภาพสมรสของบิดามารดา			/	/						/				
ภูมิลำเนาของบิดามารดา									/					
จำนวนบุตรในครอบครัว			/	/	/		/			/				
จำนวนบุตรที่ได้รับการศึกษาขั้นสูงหรือการมีพี่ศึกษาอยู่ในระดับขั้นมัธยมศึกษาหรือสูงกว่าบิดา			/	/	/									
ระยะทางจากบ้านถึงโรงเรียน			/		/									
ความสะดวกในการคมนาคม			/		/									

จากตารางที่ 11 พบว่าตัวแปรที่สัมพันธ์กับการศึกษาต่อและไม่ศึกษาต่อของนักเรียนสรุปได้ดังนี้

ความต้องการศึกษาต่อของนักเรียน จากการศึกษาของ เทียมจันทร์ จาตุกัญญาประทีป (2523) และภัทรธิรา วารินทร์ (2539) ได้ข้อค้นพบที่สอดคล้องกันว่า ความต้องการศึกษาต่อของนักเรียนเป็นตัวแปรที่สัมพันธ์กับการศึกษาต่อของนักเรียน กล่าวคือ นักเรียนที่มีความต้องการที่จะศึกษาต่อในระดับที่สูงขึ้นจะมีโอกาสศึกษาต่อมากกว่านักเรียนที่ไม่ต้องการที่จะศึกษา

ต่อในระดับที่สูงขึ้น ส่วนธวัช แก้วอนันต์ (2530) พบว่า ความต้องการศึกษาต่อของนักเรียนเป็นตัวแปรที่มีความสำคัญในการจำแนกกลุ่มของนักเรียนที่ได้ศึกษาต่อและไม่ศึกษาต่อ

ความมุ่งหวังของผู้ปกครองที่จะให้บุตรได้รับการศึกษาขั้นสูง จากการศึกษาของเทียมจันทร์ จาตุกัญญาประทีป (2523) พบว่า ความมุ่งหวังของผู้ปกครองที่จะให้บุตรได้รับการศึกษาขั้นสูง เป็นตัวแปรสำคัญในการพิจารณาและตัดสินใจว่าจะให้บุตรศึกษาต่อหรือไม่ศึกษาต่อ นอกจากนี้จากการศึกษาของธวัช แก้วอนันต์ (2530) และ ดำรง แสนสิงห์ (2534) ได้ข้อพบที่สอดคล้องกันว่า ความมุ่งหวังของผู้ปกครองที่จะให้บุตรได้รับการศึกษาขั้นสูงเป็นตัวแปรที่มีผลต่อการจำแนกกลุ่มของนักเรียนที่ได้ศึกษาต่อและไม่ศึกษาต่อ

ผลสัมฤทธิ์ทางการเรียนของนักเรียน จากการศึกษาของ จรูญ พาณิชย์ผลินไชย (2521) เทียมจันทร์ จาตุกัญญาประทีป (2523) ธวัช แก้วอนันต์ (2530) บุญเพ็ง ธานี (2530) สวลี อาชาสุวรรณ (2531) ดำรง แสนสิงห์ (2534) และ Worthington (1971) ได้ข้อค้นพบที่สอดคล้องกันว่า ระดับผลการเรียนของนักเรียนหรือผลสัมฤทธิ์ทางการเรียนของนักเรียนเป็นตัวแปรที่สัมพันธ์กับการศึกษาต่อของนักเรียน

อาชีพของบิดามารดา หรือผู้ปกครอง จากการศึกษาของ จรูญ พาณิชย์ผลินไชย (2521) วิศิษฐ์ นาจำปา (2521) อุดม ภู่ประดิษฐ์ (2521) เทียมจันทร์ จาตุกัญญาประทีป (2523) ช่อทิพย์ ราศรีเกรียงไกร (2523) สวลี อาชาสุวรรณ (2531) ดำรง แสนสิงห์ (2534) ภัทรธิรา วารินทร์ (2539) Worthigton (1971) และ Blake (1985) ได้ข้อค้นพบที่สอดคล้องกันว่า ตัวแปรอาชีพของบิดามารดาหรือผู้ปกครองเป็นตัวแปรที่สัมพันธ์กับการศึกษาต่อของนักเรียน

รายได้ของบิดามารดา หรือผู้ปกครอง จากการศึกษาของ วิศิษฐ์ นาจำปา (2521) ช่อทิพย์ ราศรีเกรียงไกร (2523) เทียมจันทร์ จาตุกัญญาประทีป (2523) ประเสริฐ แก้วเพ็ชร (2527) สวลี อาชาสุวรรณ (2531) ดำรง แสนสิงห์ (2534) Worthington (1971) และ Blake (1985) ได้ข้อค้นพบที่สอดคล้องกันว่า ตัวแปรรายได้ของบิดามารดา สัมพันธ์กับการศึกษาต่อของนักเรียน คือ บิดามารดาหรือผู้ปกครองที่มีรายได้สูง จะมีแนวโน้มให้บุตรได้รับการศึกษาต่อมากกว่าที่มีรายได้ต่ำ

ระดับการศึกษาของบิดามารดา หรือผู้ปกครอง จากการศึกษาของ จรูญ พาณิชย์ผลินไชย (2521) วิศิษฐ์ นาจำปา (2521) อุดม ภู่ประดิษฐ์ (2521) ช่อทิพย์ ราศรีเกรียงไกร (2523) เทียมจันทร์ จาตุกัญญาประทีป (2523) บุญเพ็ง ธานี (2530) สวลี อาชาสุวรรณ (2531) William และ Vimal (1968) Worthington (1971) และ Blake (1985) ได้ข้อค้นพบที่สอดคล้องกันว่า ตัวแปรระดับการศึกษาของบิดามารดา หรือผู้ปกครองเป็นตัวแปรที่สัมพันธ์กับการศึกษาต่อของนักเรียน กล่าวคือ บิดามารดา หรือผู้ปกครองที่มีการศึกษาอยู่ในระดับสูง นักเรียนจะมีแนวโน้มในการศึกษาต่อสูงด้วย

สถานภาพสมรสของบิดามารดา จากการศึกษาของ สวลี อาชาสุวรรณ (2531) ดำรง แสนสิงห์ (2534) และเทียมจันทร์ จาตุกัญญาประทีป (2523) ได้ข้อค้นพบที่สอดคล้องกันว่า ตัวแปรสถานภาพสมรสของบิดา มารดา สัมพันธ์กับการศึกษาต่อของนักเรียน กล่าวคือ นักเรียนที่บิดามารดามีสถานภาพสมรสอยู่ด้วยกัน จะมีแนวโน้มที่จะศึกษาต่อมากกว่านักเรียนที่บิดามารดามีสถานภาพสมรสหย่าร้างหรือฝ่ายใดฝ่ายหนึ่งถึงแก่กรรม

ภูมิลำเนาของบิดามารดา จากการศึกษาของ วิศิษฐ์ นาจำปา (2521) ได้ข้อค้นพบว่า ภูมิลำเนาของบิดา มารดาสัมพันธ์กับการศึกษาต่อของนักเรียน กล่าวคือ บิดามารดาหรือผู้ปกครองที่มีการย้ายถิ่นฐานหรือภูมิลำเนาเป็นประจำมีผลทำให้แนวโน้มในการศึกษต่อของนักเรียนมีน้อยกว่าบิดามารดาหรือผู้ปกครองที่ไม่มีการย้ายถิ่นฐานหรือภูมิลำเนาบ่อยครั้ง

จำนวนบุตรในครอบครัว จากการศึกษาของ เทียมจันทร์ จาตุกัญญาประทีป (2523) ประเสริฐ แก้วเพ็ชร (2527) ธวัช แก้วอนันต์ (2530) สวลี อาชาสุวรรณ (2531) และ ดำรง แสนสิงห์ (2534) ได้ข้อค้นพบที่สอดคล้องกันว่า จำนวนบุตรในครอบครัว เป็นตัวแปรที่สัมพันธ์กับการศึกษาต่อของนักเรียน

จำนวนบุตรที่ได้รับการศึกษาขั้นสูง หรือ การมีพี่ที่ศึกษาอยู่ในชั้นมัธยม หรือสูงกว่าบิดา จากการศึกษาของ เทียมจันทร์ จาตุกัญญาประทีป (2523) ธวัช แก้วอนันต์ (2530) ดำรง แสนสิงห์ (2534) และได้ข้อค้นพบที่สอดคล้องกันว่า การที่มีบุตรได้รับการศึกษาขั้นสูง หรือสูงกว่าบิดานั้น นักเรียนจะมีแนวโน้มที่จะศึกษาต่อ

ระยะทางจากบ้านถึงโรงเรียน จากการศึกษาของ ธวัช แก้วอนันต์ (2530) และ ดำรง แสนสิงห์ (2534) ได้ข้อค้นพบที่สอดคล้องกันว่า ระยะทางจากบ้านถึงโรงเรียนเป็นตัวแปรที่สัมพันธ์กับการศึกษาต่อของนักเรียน โดยนักเรียนที่มีบ้านพักอยู่ห่างไกลจากโรงเรียนที่ต้องการศึกษาต่อเป็นระยะทางมากๆ จะมีโอกาสในการศึกษาต่อน้อยกว่านักเรียนที่มีบ้านพักอยู่ใกล้โรงเรียนที่ต้องการศึกษาต่อ

ความสะดวกในการคมนาคม จากการศึกษาของ ธวัช แก้วอนันต์ (2530) และ ดำรง แสนสิงห์ (2534) ได้ข้อค้นพบที่สอดคล้องกันว่า ตัวแปรความสะดวกในการคมนาคมสัมพันธ์กับการศึกษาต่อของนักเรียน กล่าวคือ ถ้ามีความสะดวกในการคมนาคม นักเรียนจะมีแนวโน้มที่จะศึกษาต่อ มากกว่าการคมนาคมที่ไม่สะดวก

จากงานวิจัยทั้ง 14 เรื่องที่ศึกษาตัวแปรที่สัมพันธ์กับการศึกษาต่อและไม่ศึกษาต่อของนักเรียนทั้งหมดนี้ ผู้วิจัยพบว่า การวัดตัวแปรตามในการวิจัยเหล่านี้ไม่ได้วัดจากพฤติกรรมของนักเรียนที่ศึกษาต่อและไม่ศึกษาต่อโดยตรงทั้งหมด แต่วัดจากความคาดหวังหรือโอกาสในการศึกษาต่อเป็นส่วนใหญ่ ดังนั้นในการวิจัยครั้งนี้ผู้วิจัยจึงไม่ได้ศึกษาพฤติกรรมการศึกษาต่อหรือไม่ศึกษาต่อของนักเรียนโดยตรงเช่นกัน หากแต่เลือกศึกษาความคาดหวังในการศึกษาต่อแทนพฤติกรรมด้วยเหตุผล 2 ประการ

ประการแรก ความคาดหวังเป็นมโนทัศน์ที่นำมาใช้อธิบายถึงพฤติกรรมของมนุษย์มาตั้งแต่อดีตจนถึงปัจจุบัน แนวคิดพื้นฐานของกลุ่มทฤษฎีการเรียนรู้ทางสังคม (social learning theory) ได้กล่าวเอาไว้ว่า สิ่งที่ทำให้มนุษย์เกิดการเรียนรู้ได้นั้นคือ ความสัมพันธ์ระหว่างสิ่งต่างๆ ไม่ว่าจะเป็นการเรียนรู้ที่เกิดจากความสัมพันธ์ระหว่างเหตุการณ์กับเหตุการณ์ หรือการเรียนรู้ที่เกิดจากความสัมพันธ์ระหว่างพฤติกรรมกับผลที่เกิดขึ้นจากพฤติกรรมนั้น มนุษย์จะเกิดการเรียนรู้ว่า เมื่อมีเหตุการณ์หนึ่งเกิดขึ้นแล้วจะต้องมีเหตุการณ์หนึ่งเกิดตามขึ้นมา ซึ่งสิ่งนั้นก็คือความคาดหวัง และความคาดหวังนี้จะทำให้มนุษย์ตัดสินใจทำหรือไม่ทำพฤติกรรมหนึ่งเพื่อให้เกิดผลตามที่ปรารถนา ดังนั้นจึงอาจกล่าวได้ว่า สิ่งที่กำหนดพฤติกรรมของมนุษย์คือความรู้สึกนึกคิดที่เป็นความคาดหวังและการตัดสินใจกระทำตามความคาดหวัง และพฤติกรรมของมนุษย์เป็นผลของความคาดหวัง (ภิญโญ วรรณสุข, 2540; Bandura, 1978; Greenough, Blake and Wallace, 1987; David and Newstorm, 1985) ผู้วิจัยจึงเชื่อว่า พฤติกรรมที่จะศึกษาต่อหรือไม่ศึกษาต่อของนักเรียนหลังจากที่จบการศึกษาขั้นพื้นฐานนั้นเป็นผลมาจากความคาดหวัง

จากทฤษฎีทางจิตวิทยาที่ว่า พฤติกรรมเป็นผลจากการที่คนเราเลือกปฏิกิริยาตอบสนองกับสิ่งเร้าตามกระบวนการจูงใจ (motivation process) ประกอบด้วย ความต้องการและแรงขับ (needs and drives) หรือแรงขับและแรงจูงใจ (drives and motives) เพื่อนำไปสู่การแสดงออกทางพฤติกรรมที่มีเป้าหมาย (motivated behavior) และที่เป็นเป้าหมาย (goals) โดยที่กระบวนการจูงใจเป็นผลมาจากความคาดหวัง (expectation) นั่นเอง (ภิญโญ วรรณสุข, 2540; Bandura, 1969, 1977, 1986; Luthan, 1985, 1988) ดังนั้น การศึกษาความคาดหวังย่อมได้ผลเช่นเดียวกันหรือใกล้เคียงกับการศึกษาพฤติกรรมโดยตรง

เหตุผลประการที่สองที่ผู้วิจัยเลือกศึกษาความคาดหวังในการศึกษาต่อของนักเรียน เนื่องจากความเชื่อที่ว่า หากได้มีการศึกษาและทราบความคาดหวังในการศึกษาของนักเรียน ผู้บริหาร คณะครูและผู้ที่มีส่วนเกี่ยวข้องกับนักเรียนจะมีโอกาสเปลี่ยนความคาดหวังของนักเรียนจากไม่ศึกษาต่อเป็นศึกษาต่อได้ทันเหตุการณ์กว่าการวิจัยที่ศึกษาพฤติกรรมโดยตรง

จะเห็นได้ว่าจากงานวิจัยที่ได้ศึกษามาทั้ง 14 เรื่องนั้น มีตัวแปรที่งานวิจัย 10 เรื่องเห็นสอดคล้องกันอยู่ 2 ตัวแปร คือ ตัวแปรระดับการศึกษาของบิดามารดาหรือผู้ปกครอง และตัวแปรอาชีพของบิดามารดาหรือผู้ปกครอง ตัวแปรที่งานวิจัย 8 เรื่อง เห็นสอดคล้องกัน คือตัวแปรรายได้ของบิดามารดาหรือผู้ปกครอง ตัวแปรที่งานวิจัยเห็นสอดคล้องกัน 7 เรื่อง คือตัวแปรผลสัมฤทธิ์ทางการเรียนของนักเรียน และตัวแปรที่งานวิจัยเห็นสอดคล้องกัน 5 เรื่อง คือ ตัวแปรจำนวนบุตรในครอบครัว ผู้วิจัยจึงสนใจที่จะนำตัวแปรทั้ง 5 ตัวนี้มาศึกษาหาความสัมพันธ์กับความคาดหวังในการศึกษาต่อของนักเรียนชั้นมัธยมศึกษาปีที่ 3 โรงเรียนขยายโอกาสทางการศึกษาขั้นพื้นฐาน ดังนี้

อาชีพของผู้ปกครอง

อาชีพของผู้ปกครองเป็นตัวแปรสำคัญที่มีอิทธิพลต่อการศึกษาต่อในระดับที่สูงขึ้นของนักเรียน เนื่องจากการศึกษาเป็นรากฐานสำคัญในการประกอบอาชีพของบุคคล บุคคลที่จบการศึกษาในระดับสูงมักจะมีโอกาสในการประกอบอาชีพที่ดีกว่าบุคคลที่จบการศึกษาในระดับการศึกษาที่ต่ำกว่า ดังนั้นผู้ที่ประกอบอาชีพที่ดีจึงมักจะเห็นความสำคัญของการศึกษาเป็นอย่างยิ่ง จากการศึกษาของ ดำรง แสนสิงห์ (2534) เทียมจันทร์ จาตุกัญญาประทีป (2523) บุญเพ็ง ธานี (2530) และสวลี อาชาสุวรรณ (2532) ได้ข้อค้นพบว่า อาชีพของผู้ปกครองจะส่งผลต่อโอกาสในการศึกษาต่อของนักเรียนที่สำเร็จการศึกษาซึ่งสอดคล้องกับ Blake (1985) Waite, Rindfuss and tray (1986) และ Worthington and Grant (1971) ที่พบว่า อาชีพของผู้ปกครอง เป็นตัวแปรที่สัมพันธ์กับการศึกษาต่อของนักเรียน กล่าวคือผู้ปกครองของนักเรียนที่ประกอบอาชีพที่ดีมีรายได้สูงมีผลต่อโอกาสที่บุตรจะได้รับการศึกษาขั้นสูงมากยิ่งขึ้น

ระดับการศึกษาของผู้ปกครอง

ระดับการศึกษาของผู้ปกครองเป็นดัชนีที่สะท้อนให้เห็นถึงประสบการณ์และการเห็นความสำคัญของการศึกษาได้เช่นกัน กล่าวคือ ผู้ปกครองที่มีระดับการศึกษาสูงมักจะเห็นความสำคัญของการศึกษาของบุตร ดังรายงานการวิจัยของ Jamison and Lockheed (1987) และ Rehberg and Westby (1967) ที่พบว่าการสำเร็จการศึกษาและระดับการศึกษาของบิดาเป็นปัจจัยที่กำหนดการศึกษาต่อของบุตรสอดคล้องกับงานวิจัยของ Chemichovsky (1985) เช่นกันว่า ระดับการศึกษาของผู้ปกครองมีผลต่อการได้รับการศึกษาของบุตรมากที่สุด ส่วน Waite, Rindfuss and Tray (1986) ได้ศึกษาถึงความมุ่งหวังของมารดาเกี่ยวกับการศึกษาของบุตรพบว่าระดับการศึกษาของมารดามีผลต่อความมุ่งหวังทางการศึกษาของมารดาที่จะให้บุตรได้รับการศึกษาขั้นสูง

รายได้ของผู้ปกครอง

การศึกษาขั้นพื้นฐานนั้นเป็นการจัดแบบให้เปล่าจากรัฐ แต่การศึกษาต่อในระดับที่สูงขึ้นนั้นผู้ปกครองของนักเรียนต้องรับผิดชอบภาระค่าใช้จ่ายในการศึกษาเองทั้งหมด ซึ่งผู้ปกครองเป็นจำนวนมากไม่สามารถรับผิดชอบค่าใช้จ่ายในส่วนนี้ได้ ทำให้นักเรียนที่สำเร็จการศึกษาในระดับขั้นพื้นฐาน ไม่มีโอกาสได้ศึกษาต่อในระดับที่สูงขึ้น จึงกล่าวได้ว่า รายได้ของผู้ปกครองมีความสำคัญต่อการศึกษาต่อเป็นอย่างยิ่ง จากการศึกษาของเทียมจันทร์ จาตุกัญญาประทีป (2523) พบว่า รายได้ของผู้ปกครองเป็นตัวแปรตัวหนึ่งที่ทำให้นักเรียนมีโอกาสได้ศึกษาต่อและไม่ได้ศึกษาต่อ โดยนักเรียนที่ได้ศึกษาต่อนั้นผู้ปกครองจะมีรายได้สูงกว่านักเรียนที่ไม่ได้ศึกษาต่อซึ่งสอดคล้องกับผลการวิจัยของ บุญเพ็ง ธานี (2530) และ ประเสริฐ แก้วเพชร (2528)

จำนวนบุตรในครอบครัว

จำนวนบุตรในครอบครัวเป็นตัวแปรที่สำคัญตัวหนึ่งในการได้เข้าศึกษาต่อในระดับการศึกษาที่สูงขึ้นของนักเรียน กล่าวคือ หากผู้ปกครองหรือครอบครัวใดมีจำนวนบุตรในครอบครัวเป็นจำนวนมาก จำนวนบุตรในครอบครัวมักจะได้รับการศึกษาในระดับที่แตกต่างกันออกไป และ

หากผู้ปกครองหรือครอบครัวใดมีจำนวนบุตรในครอบครัวเป็นจำนวนมากบุตรจะมีโอกาสได้ศึกษาต่อน้อยกว่าผู้ปกครองหรือครอบครัวใดมีจำนวนบุตรในครอบครัวจำนวนน้อย ซึ่งสอดคล้องกับงานวิจัยของเทียมจันทร์ จาตุกัญญาประทีป (2523) ประเสริฐ แก้วเพ็ชร (2527) ธวัช แก้วอนันต์ (2530) สวลี อาชาสุวรรณ (2531) และ ดำรง แสนสิงห์ (2534) ที่ได้ข้อค้นพบที่สอดคล้องกันว่า จำนวนบุตรในครอบครัว เป็นตัวแปรที่สัมพันธ์กับการศึกษาต่อของนักเรียน

ผลสัมฤทธิ์ทางการเรียน

ผลสัมฤทธิ์ทางการเรียนของนักเรียนเป็นดัชนีที่บ่งชี้ถึงความสามารถและศักยภาพในการเรียนของนักเรียนที่สำคัญยิ่งอย่างหนึ่งที่จะใช้ประกอบการตัดสินใจในเรื่องความคาดหวังในการศึกษาต่อของนักเรียน จากการศึกษาของ จรูญ พานิชผลินไชย (2521) ดำรง แสนสิงห์ (2534) เทียมจันทร์ จาตุกัญญาประทีป (2523) ธวัช แก้วอนันต์ (2533) บุญเพ็ง ธานี (2530) ประเสริฐ แก้วเพ็ชร (2528) สวลี อาชาสุวรรณ (2532) และ Worthington and Grant (1971) สอดคล้องกันว่า นักเรียนที่มีผลสัมฤทธิ์ทางการเรียนในระดับการศึกษาขั้นพื้นฐานดีจะมีโอกาสได้ศึกษาต่อในระดับการศึกษาสูงกว่านักเรียนที่มีผลสัมฤทธิ์ทางการเรียน ปานกลาง หรือ ต่ำ นอกจากนี้ เทียมจันทร์ จาตุกัญญาประทีป (2523) ยังพบอีกว่า นักเรียนที่มีผลสัมฤทธิ์ทางการเรียนสูงจะมีความมุ่งหวังในการศึกษาต่อสูงกว่านักเรียนที่มีผลสัมฤทธิ์ทางการเรียนต่ำ

สำหรับตัวแปรทั้ง 5 ตัวที่นำมาศึกษาในครั้งนี้ มีตัวแปร 3 ตัว ประกอบด้วยตัวแปรรายได้ของบิดา มารดา หรือผู้ปกครอง ตัวแปรผลสัมฤทธิ์ทางการเรียนของนักเรียน ตัวแปรจำนวนบุตรในครอบครัว เป็นตัวแปรที่อยู่ในมาตรอันตรภาคขึ้นไป ส่วนตัวแปรอีก 2 ตัวแปร คือ ตัวแปรระดับการศึกษาของบิดา มารดา หรือผู้ปกครอง ตัวแปรอาชีพของบิดามารดา เป็นตัวแปรที่อยู่ในมาตรอันดับและมาตรนามบัญญัติ แต่ผู้วิจัยได้นำตัวแปรทั้งหมดนี้มาวิเคราะห์โดยทำให้ตัวแปรทั้งหมดอยู่ในมาตรอันดับ ทั้งนี้เพื่อประโยชน์ในการเก็บรวบรวมข้อมูล กล่าวคือจะมีความสะดวกในการตอบแบบสอบมากขึ้นเนื่องจากผู้ตอบจะสามารถเลือกคำตอบได้จากข้อความที่ได้จัดเตรียมไว้ให้ และรวมทั้งทำให้การใช้เวลาในการตอบแบบสอบถามรวดเร็วขึ้น ผู้ตอบจะสามารถตอบได้ง่ายและมีความต้องการที่จะตอบแบบสอบถามมากขึ้น

ตอนที่ 4 การนำการวิเคราะห์ล็อกลิเนียร์มาใช้ในประเทศไทย

จากการศึกษางานวิจัย พบว่า มีงานวิจัยของชื่นชม เจริญยุทธ (2529) อ้างถึงในสุชาติ ประสิทธิ์รัฐสินธุ์ และกรรณิการ์ สุขเกษม (2533) ที่ศึกษาปัจจัยที่ส่งผลต่อระยะเวลาการเลี้ยงบุตรด้วยนมมารดาของสตรีครรภ์แรกที่ทำงานนอกบ้านในเขตกรุงเทพมหานคร โดยกำหนดให้ระยะเวลาการเลี้ยงบุตรด้วยนมมารดาเป็นตัวแปรตาม 2 กลุ่ม ตัวแปรอิสระมี 8 ตัวแปร แต่ละตัวแปรได้จัดแบ่งเป็น 2 กลุ่ม แยกวิเคราะห์ตัวแปรที่ส่งผลต่อตัวแปรตามครั้งละ 2 ตัวแปร ด้วยตารางการณ์จร 3 มิติ

และจากงานวิจัยของทวีพร บุญวาณิช (2541) ได้ประยุกต์ใช้โมเดลล็อกลิเนียร์ในการวิเคราะห์สาเหตุเพื่อศึกษาปัจจัยที่ส่งผลต่อระยะเวลาในการศึกษาของมหาบัณฑิตทางสังคมศาสตร์จุฬาลงกรณ์มหาวิทยาลัย โดยกำหนดให้ระยะเวลาในการทำวิทยานิพนธ์เป็นตัวแปรตาม และ ตัวแปรต้นมี 6 ตัวแปร ซึ่งตัวแปรที่นำมาศึกษาทั้งหมดเป็นตัวแปรมาตรนามบัญญัติ

จากการศึกษางานวิจัยที่ใช้ล็อกลิเนียร์ในประเทศไทย พบว่า ยังไม่มีการนำตัวแปรมาตรอันดับมาใช้ในการวิจัย ผู้วิจัยจึงสนใจที่จะนำตัวแปรอันดับมาใช้ในงานวิจัยครั้งนี้

สมมุติฐานการวิจัย

จากการศึกษางานวิจัยที่เกี่ยวข้องกับการศึกษาต่อของนักเรียนหลายๆเรื่องดังรายละเอียดในบทที่ 2 วรรณคดีที่เกี่ยวข้อง ตอนที่ 3 ผู้วิจัยจึงตั้งสมมุติฐานในการวิจัยครั้งนี้ว่า ตัวแปรอิสระทุกตัวที่ผู้วิจัยศึกษาค้นคว้า ได้แก่ ระดับการศึกษาของผู้ปกครอง อาชีพของผู้ปกครอง รายได้ของผู้ปกครอง ตัวแปรสัมฤทธิ์ทางการเรียน และจำนวนบุตร จะมีอิทธิพลหลัก (main effect) ต่อตัวแปรตามคือ ความคาดหวังในการศึกษาต่อของนักเรียนชั้นมัธยมศึกษาปีที่ 3 โรงเรียนขยายโอกาสทางการศึกษาขั้นพื้นฐาน

สำหรับกรณีของอิทธิพลปฏิสัมพันธ์ (interaction effect) นั้น เนื่องจากผู้วิจัยไม่พบว่ามีแนวคิดหรือทฤษฎีที่ชัดเจนเกี่ยวกับการที่จะมีเทอมอิทธิพลปฏิสัมพันธ์เกิดขึ้น ผู้วิจัยจึงไม่ได้ตั้งสมมุติฐานเกี่ยวกับเทอมอิทธิพลปฏิสัมพันธ์เอาไว้