บทที่ 2

ความรู้พื้นฐานและงานวิจัยที่เกี่ยวข้อง

ในบทนี้จะกล่าวถึง ความรู้พื้นฐานและงานวิจัยที่เกี่ยวข้องกับเทคนิคการวิเคราะห์ข้อมูลเชิงกลุ่ม (Analysis of Categorical Data) ในรูปของตารางการณ์จร (Contingency Table) โดยใช้ตัวแบบล็อกลิเนียร์ (Log-linear Model) ตลอดจนงานวิจัยที่เกี่ยวกับปัจจัยที่มีผลต่อการสูญเสียการได้ยินของคนงานในโรงงานอุตสาหกรรม เพื่อเป็นแนวทางในการศึกษาและทำความเข้าใจเกี่ยวกับตัวแบบล็อกลิเนียร์มากยิ่งขึ้น ซึ่งมีรายละเอียดดังนี้

2.1 มาตรวัดสำหรับตัวแปรเชิงกลุ่ม

2.2 แผนการสุ่มตัวอย่างในการได้มาซึ่งข้อมูล (Sampling schemes)

2.3 ตารางการณ์จรสองทาง (Two-Way Contingency Table)

2.4 อัตราส่วนออดส์ (Odds Ratio)

2.5 ฟังก์ชันภาวะน่าจะเป็นและตัวประมาณค่าภาวะน่าจะเป็นสูงสุด (Likelihood Function and Maximum Likelihood Estimator or MLE)

2.6 การทดสอบภาวะรูปดีแบบต่างๆ

  2.6.1 การทดสอบความเป็นอิสระโดยใช้การทดสอบไคสแควร์ (Chi-Square Test)

  2.6.2 ตัวสถิติอัตราส่วนภาวะน่าจะเป็นไคสแควร์ (Likelihood-Ratio Chi-Squared Statistic)

2.7 ตัวแบบเชิงเส้นที่วางนัยทั่วไป (Generalized Linear Model or GLM)

2.8 งานวิจัยที่เกี่ยวข้อง

2.1 มาตรวัดสำหรับตัวแปรเชิงกลุ่ม

1) มาตรวัดนามบัญญัติ (Nominal Scale) เป็นมาตรการวัดที่มีลักษณะหยาบที่สุด เป็นการวัดในลักษณะของการจำแนกสิ่งต่างๆ ออกเป็นประเภท ชนิดหรือกลุ่ม ซึ่งจะมีการกำหนดตัวเลขแทนประเภท ชนิดหรือกลุ่มนั้นๆ โดยแต่ละกลุ่มไม่อาจจัดเรียงลำดับความมาก-น้อยได้ ค่าตัวเลขที่กำหนดให้สำหรับแต่ละกลุ่มจะไม่มีความหมายในการคำนวณ แต่เป็นการบอกถึงความแตกต่างเท่านั้น ไม่สามารถบอกได้ว่ากลุ่มใดดีกว่ากัน เช่น เพศ (ชาย/หญิง) สถานภาพสมรส (โสด / สมรส / หม้าย / หย่าร้าง) การกำหนดเบอร์ให้กับนักกีฬา เป็นต้น

2) มาตรวัดเรียงลำดับ (Ordinal Scale) เป็นมาตรวัดที่มีรายละเอียดเพิ่มขึ้นมากกว่ามาตรวัดนามบัญญัติ ตัวเลขที่กำหนดให้กับข้อมูลสามารถบอกอันดับและทิศทางความแตกต่างได้ นั่นคือ สามารถบอกได้ว่ากลุ่มใดดีกว่ากัน แต่ไม่สามารถบอกปริมาณหรือขนาดความมาก-น้อยได้ว่าเป็นเท่าใด เช่น ระดับการศึกษา (ปริญญาตรี/ปริญญาโท /ปริญญาเอก) ประสิทธิภาพของเครื่องจักร (ดี/ปานกลาง/แย่) เป็นต้น

3) มาตรวัดอันตรภาค (Interval Scale) เป็นมาตรวัดที่มีรายละเอียดเพิ่มขึ้นมากกว่ามาตรวัดเรียงลำดับ โดยข้อมูลจะมีการแบ่งออกเป็นช่วงๆ เรียงตามลำดับและแต่ละช่วงมีขนาดเท่าๆ กัน ซึ่งมีคุณสมบัติที่สำคัญคือไม่มีศูนย์ที่แท้จริง เช่นอุณหภูมิ (อุณหภูมิ 0°C ไม่ได้หมายความว่าไม่มีความร้อน) คะแนนสอบ (คนที่สอบได้ 0 คะแนน ไม่ได้หมายความว่าไม่มีความรู้เลย) เป็นต้น จะมีจุดเริ่มต้นที่เรียกว่า ศูนย์สมมติ (Arbitrary Zero or Relative Zero) มาตรวัดนี้นอกจากจะแบ่งเป็นกลุ่มและเรียงลำดับแล้ว ยังสามารถคำนวณเพื่อบอกปริมาณความแตกต่างระหว่างกลุ่มได้ แต่ไม่สามารถระบุถึงสัดส่วนแสดงความเป็นจำนวนกี่เท่าตัวได้ เช่น นักเรียนที่สอบได้คะแนน 40 คะแนน ไม่สามารถระบุได้ว่ามีความรู้เป็น 2 เท่าของนักเรียนที่สอบได้คะแนน 20 คะแนน เป็นต้น

2.2 แผนการสุ่มตัวอย่างในการได้มาซึ่งข้อมูล (Sampling schemes)

ข้อมูลเชิงกลุ่มในตารางการณ์จร มีที่มาจากแผนการสุ่มตัวอย่าง ดังนี้

1) แผนการสุ่มตัวอย่างแบบปัวส์ซง (Poisson sampling scheme) เป็นการเก็บรวบรวมข้อมูลโดยใช้กระบวนการปัวส์ซง (Poisson processes) กล่าวคือ ทำการเก็บรวบรวมข้อมูลเกี่ยวกับปัจจัยกลุ่มใดกลุ่มหนึ่ง ภายในช่วงระยะเวลาใดเวลาหนึ่งที่กำหนด โดยไม่ทราบถึงจำนวนหน่วยตัวอย่างในกลุ่มนั้นล่วงหน้า เมื่อเสร็จสิ้นกระบวนการปัวส์ซง 1 กระบวนการ จึงได้จำนวนหน่วยตัวอย่างในแต่ละกลุ่มรวมถึงจำนวนหน่วยตัวอย่างรวมทั้งหมด

2) แผนการสุ่มตัวอย่างแบบพหุนาม (Multinomial sampling scheme) เป็นการเก็บรวบรวมข้อมูลโดยกำหนดขนาดตัวอย่างทั้งหมด (n) ไว้ล่วงหน้า โดยไม่คำนึงถึงระยะเวลาในการเก็บรวบรวมข้อมูล และเมื่อเก็บรวมรวมข้อมูลเสร็จสิ้นจึงได้จำนวนหน่วยตัวอย่างในแต่ละกลุ่ม

3) แผนการสุ่มตัวอย่างแบบพหุนามผลคูณ (Product-Multinomial sampling scheme) เป็นการสุ่มพหุนามหลายขั้นตอน กล่าวคือ ทำการเก็บรวบรวมข้อมูลเป็นลำดับทีละปัจจัย โดยกำหนดขนาดตัวอย่างและทำการสุ่มตัวอย่างเพื่อเก็บรวบรวมข้อมูลในปัจจัยแรกเสียก่อน แล้วจึงสุ่มตัวอย่างแบบพหุนามจากข้อมูลปัจจัยแรกที่เก็บรวบรวมได้ เพื่อให้ได้ข้อมูลในปัจจัยที่สองตรงกัน (cross-classified) กับปัจจัยแรก หากมีปัจจัยที่สามก็ทำการสุ่มตัวอย่างแบบพหุนามจากข้อมูลสองปัจจัยแรกตรงกัน ทำเช่นนี้ไปเรื่อยๆ จนครบทุกปัจจัยที่ต้องการศึกษา

2.3 ตารางการณ์จรสองทาง (Two–Way Contingency Table)

การที่สมาชิกของตัวอย่างที่สุ่มจากประชากรถูกจำแนกออกเป็นกลุ่มตัวแปรตั้งแต่ 2 กลุ่มขึ้นไปแล้วนำมาเขียนอยู่ในรูปของตารางและทำการวิเคราะห์ตัวอย่างนั้น โดยใช้ความถี่ของค่าสังเกตซึ่งได้มาจากการนับที่เกิดขึ้นร่วมกันในแต่ละระดับของตัวแปร โดยมีข้อตกลงว่า หน่วยตัวอย่างแต่ละหน่วยต้องถูกจำแนกลงในแต่ละช่อง (Cell) ได้อย่างอิสระ (Independent) เนื่องจากลักษณะของตัวแปรถูกแบ่งออกเป็นกลุ่มๆ ที่ไม่ต่อเนื่อง เราจะเรียกตารางได้ว่า ตารางการณ์จร (Contingency table or Cross-classification table) โดยตารางจะมีขนาด (Dimension) หรือจำนวนช่องเป็นเท่าไรนั้น ขึ้นอยู่กับจำนวนตัวแปรที่สนใจศึกษา

ตารางการณ์จรสองทาง คือ ตารางแจกแจงความถี่ของข้อมูลหรือค่าสังเกตตามคุณลักษณะหรือตัวแปรเชิงกลุ่มที่สนใจจำนวน 2 ตัวแปร สมมติว่าตัวแปรที่สนใจศึกษาคือตัวแปร X ซึ่งแบ่งออกเป็น r กลุ่ม และตัวแปร Y ซึ่งแบ่งออกเป็น c กลุ่ม โดยแต่ละตัวแปรอาจมีมาตรวัดนามบัญญัติ มาตรวัดเรียงลำดับหรือมาตรวัดอันตรภาคก็ได้ และหากจัดหมวดหมู่ (Combination) ตามลักษณะของตัวแปรจะสามารถจำแนกได้ rc กลุ่ม ซึ่งแสดงถึงขนาดของตารางการณ์จร (ขนาด $r \times c$) หรือจำนวนช่อง (Cells) ทั้งหมด โดยที่ “ช่อง” หมายถึงตำแหน่งที่แสดงค่าความถี่ในตารางการณ์จร

ให้ X_{ij} เป็นตัวแปรสุ่มซึ่งใช้แทนความถี่ (Frequency) ของจำนวนนับหรือจำนวนหน่วยตัวอย่างที่สอดคล้องกับกลุ่มที่ i ของตัวแปร X เมื่อ $i = 1, 2, ..., r$ และกลุ่มที่ j ของตัวแปร Y เมื่อ $j = 1, 2, ..., c$ หรือเป็นจำนวนหน่วยตัวอย่างที่ตกอยู่ในเซลล์ที่ ij โดยสัญลักษณ์ ij แทนตำแหน่งของช่องในตารางการณ์จร ดังแสดงในตาราง 2.1

ตาราง 2.1 ลักษณะทั่วไปของตารางการณ์จรสองทาง ขนาด $r \times c$ ที่ประกอบด้วยความถี่ของจำนวนนับในแต่ละช่อง

ตัวแปร X	ตัวแปร Y				รวม
	Y_1	Y_2	…	Y_c	
X_1	x_{11}	x_{12}	…	x_{1c}	$n_{1\cdot}$
X_2	x_{21}	x_{22}	…	x_{1c}	$n_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$	x_{ij}	$\vdots$	$\vdots$
X_r	x_{r1}	x_{r2}	…	x_{rc}	$n_{r\cdot}$
รวม	$n_{\cdot 1}$	$n_{\cdot 2}$	…	$n_{\cdot c}$	$n_{\cdot\cdot} = n$

เมื่อ $n_{i\cdot} = \sum_{j=1}^{c} x_{ij}$, $n_{\cdot j} = \sum_{i=1}^{r} x_{ij}$ และ $\sum_{j=1}^{c} n_{i\cdot} = \sum_{i=1}^{r} n_{\cdot j} = \sum_{i=1}^{r}\sum_{j=1}^{c} x_{ij} = n.. = n$

ให้ p_{ij} แทนความน่าจะเป็นที่ตัวแปรสุ่ม X_{ij} จะตกอยู่ในเซลล์ ij หรือความน่าจะเป็นร่วมของการเกิดตัวแปร X ในกลุ่ม (แถว) ที่ i และการเกิดตัวแปร Y ในกลุ่ม (สดมภ์) ที่ j ซึ่งเป็นกลุ่มที่ไขว้ตรงกัน (Cross-classified) ทั้งนี้ p_{ij} ก็สามารถแสดงในตารางการณ์จรเช่นเดียวกับ X_{ij} แต่มีข้อแตกต่างบางประการ ได้แก่

$$p_{i\cdot} = \sum_{j=1}^{c} p_{ij},\ p_{\cdot j} = \sum_{i=1}^{r} p_{ij} \text{ และ } \sum_{j=1}^{c} p_{i\cdot} = \sum_{i=1}^{r} p_{\cdot j} = \sum_{i=1}^{r}\sum_{j=1}^{c} p_{ij} = p_{\cdot\cdot} = 1$$

เมื่อ p_{ij} แทนความน่าจะเป็นร่วม (Joint Probability)

$p_{i\cdot}$ แทนความน่าจะเป็นชายขอบ (Marginal Probability) ทางแถว

$p_{\cdot j}$ แทนความน่าจะเป็นชายขอบ (Marginal Probability) ทางสดมภ์

ในตารางการณ์จรสองทางนั้น ถ้ากำหนดให้ตัวแปรหนึ่งเป็นตัวแปรตาม (Dependent variable) และอีกตัวแปรหนึ่งเป็นตัวแปรอิสระ (Independent variable) ซึ่งไม่เป็นตัวแปรสุ่ม จะทำให้การแจกแจงร่วมของตัวแปรทั้งสองไม่มีความหมาย แต่เมื่อกำหนดให้ตัวแปรทั้งสองตัวเป็นตัวแปรตาม จะสามารถใช้การแจกแจงความน่าจะเป็นร่วมของตัวแปรทั้งสอง และความน่าจะเป็นแบบมีเงื่อนไข (Condition probability) ของตัวแปร X เมื่อกำหนดตัวแปร Y มาช่วยในการอธิบายความสัมพันธ์ของตัวแปรทั้งคู่ (ถ้าให้ความสำคัญกับตัวแปร X เมื่อตัวแปร Y มีการเปลี่ยนแปลง จะทำให้ทราบว่าค่าของตัวแปร X มีการเปลี่ยนแปลงเป็นอย่างไร) โดยเราสามารถหาความน่าจะเป็นแบบมีเงื่อนไข ($p_{j|i}$) ได้จาก

$$p_{j|i} = \frac{p_{ij}}{p_{i\cdot}}$$

แต่ถ้าตัวแปรทั้งสองเป็นอิสระต่อกันแล้ว จะได้ว่า

$$p_{j|i} = p_{i\cdot}p_{\cdot j}$$

จากการแจกแจงความน่าจะเป็นร่วม การแจกแจงความน่าจะเป็นแบบชายขอบและการแจกแจงแบบมีเงื่อนไข สามารถแสดงการณ์จรสองทาง ได้ดังตาราง 2.2

ตาราง 2.2 ความน่าจะเป็นร่วม ความน่าจะเป็นชายขอบและความน่าจะเป็นแบบมีเงื่อนไข

ตัวแปร X	ตัวแปร Y				รวม
	Y_1	Y_2	…	Y_c	
X_1	p_{11}	p_{12}	…	p_{1c}	$p_{1\cdot}$
	$(p_{1\|1})$	$(p_{2\|1})$	…	$(p_{c\|1})$	(1.0)
X_2	p_{21}	p_{22}	…	p_{2c}	$p_{2\cdot}$
	$(p_{1\|2})$	$(p_{2\|2})$	…	$(p_{c\|2})$	(1.0)
$\vdots$	$\vdots$	$\vdots$	p_{ij}	$\vdots$	$\vdots$
X_r	p_{r1}	p_{r2}	…	p_{rc}	$p_{r\cdot}$
	$(p_{1\|r})$	$(p_{2\|r})$	…	$(p_{c\|r})$	(1.0)
รวม	$p_{\cdot 1}$	$p_{\cdot 2}$	…	$p_{\cdot c}$	1.0

2.4 อัตราส่วนออดส์ (Odds Ratio)

จากตาราง 2.2 สมมติว่าตัวแปรแต่ละตัวมี 2 ระดับ ตารางการณ์จรที่ได้จะมีขนาด 2×2 จะได้ว่าในแถวที่ 1 ค่าออดส์ของ Y_1 หมายถึงการเปรียบเทียบแนวโน้มการเกิดของระดับหรือกลุ่ม Y_1 เมื่อเปรียบเทียบกับระดับหรือกลุ่ม Y_2 จะแทนด้วยสัญลักษณ์ $\varphi_1 = P_{1|1}/P_{2|1}$ ในทำนองเดียวกันค่าออดส์ของ Y_1 ในแถวที่ 2 จะแทนด้วย $\varphi_2 = P_{1|2}/P_{2|2}$ และสำหรับการแจกแจงร่วมออดส์ของ Y_1 ภายในแถวที่ i จะแทนด้วย $\varphi_i = P_{i1}/P_{i2}$, $i = 1,2$ ค่าของ φ_i แต่ละค่าจะมีค่าเป็นบวกและจะมีค่ามากกว่า 1 เมื่อระดับหรือกลุ่ม Y_1 มีแนวโน้มการเกิดมากกว่าระดับหรือกลุ่ม Y_2 เช่น $\varphi_1 = 3$ หมายความว่าในแถวที่ 1 นั้น ระดับหรือกลุ่ม Y_1 มีแนวโน้มการเกิดมากเป็น 3 เท่าของระดับหรือกลุ่ม Y_2 สำหรับการเปรียบเทียบค่าออดส์ทั้งสอง ทำได้โดยการเปรียบเทียบกันในรูปอัตราส่วน ดังนี้

$$\theta = \frac{\varphi_1}{\varphi_2}$$

เราจะเรียกอัตราส่วนข้างต้นว่า อัตราส่วนออดส์ ซึ่งสามารถแสดงในรูปแบบของการแจกแจงความน่าจะเป็น ได้ดังนี้

$$\theta = \frac{P_{11}/P_{12}}{P_{21}/P_{22}} = \frac{P_{11}P_{22}}{P_{12}P_{21}}$$

อัตราส่วนออดส์ ยังสามารถเรียกได้อีกอย่างว่า Cross-Product Ratio เพราะเป็นอัตราส่วนของ $P_{11}P_{22}$ และ $P_{12}P_{21}$ ซึ่งเป็นความน่าจะเป็นจากช่องตรงกันข้ามทางด้านทะแยงมุมของตารางการณ์จรขนาด 2 × 2 เช่น $\theta = 5$ หมายความว่า ออดส์ของ Y_1 ในแถวที่ 1 สูงกว่าออดส์ของ Y_1 ในแถวที่ 2 เป็น 5 เท่า หรือกล่าวอีกนัยหนึ่งว่า หน่วยในแถวที่ 1 มีแนวโน้มเป็นระดับหรือกลุ่มของ Y_1 มากกว่าหน่วยในแถวที่ 2 ถึง 5 เท่า

โดยทั่วไป จะนิยมใช้ค่า $\ln(\theta)$ โดยความเป็นอิสระของตัวแปรนั้นจะสมมูลกับ $\ln(\theta) = 0$ ค่า $\ln(\theta)$ นี้เป็นค่าสมมาตร ซึ่งการสลับกันของแถวหรือสดมภ์ จะทำให้ค่า $\ln(\theta)$ เปลี่ยนแปลงเฉพาะเครื่องหมายเท่านั้น นั่นคือค่า $\ln(\theta)$ จะเท่ากัน 2 ค่า ยกเว้นเครื่องหมายจะต่างกัน เช่น $\ln(3) = 1.10$ และ $\ln(\frac{1}{3}) = -1.10$

คุณสมบัติของอัตราส่วนออดส์

1. เมื่อความน่าจะเป็นของช่องทุกช่องในตารางการณ์จรมีค่ามากกว่าศูนย์ จะทำให้อัตราส่วนออดส์ที่ได้มีค่าเป็นบวกและเมื่อความน่าจะเป็นดังกล่าวมีค่าเท่ากับศูนย์ จะทำให้อัตราส่วนออดส์ที่ได้มีค่าเป็น 0 หรือ ∞
2. ตัวแปรจะเป็นอิสระต่อกันเมื่อ $\theta = 1$ หรือ $P_{1|1} = P_{1|2}$
3. ถ้า $1 < \theta < \infty$ หมายความว่า $P_{1|1} > P_{1|2}$ และถ้า $0 < \theta < 1$ หมายความว่า $P_{1|1} < P_{1|2}$
4. ค่าของอัตราส่วนออดส์จะไม่เปลี่ยนแปลงเมื่อตารางการณ์จรสองทางสลับด้านกัน นั่นคือแถวกลายเป็นสดมภ์ และสดมภ์กลายเป็นแถว

สำหรับข้อมูลตัวอย่างซึ่งแสดงอยู่ในรูปความถี่ (n_{ij}) ของตารางณ์การจรสองทาง สามารถประมาณอัตราส่วนออดส์ (θ) ได้ดังนี้

$$\hat{\theta} = \frac{n_{11}n_{22}}{n_{12}n_{21}}$$

สำหรับตารางการณ์จรขนาด $r \times c$ การคำนวณอัตราส่วนออดส์ สามารถใช้ช่อง 4 ช่องในลักษณะสี่เหลี่ยมพื้นผ้าใดๆ มาหาอัตราส่วนออดส์ได้ทั้งหมด $\binom{r}{2}\binom{c}{2}$ ค่า เช่นในแถวที่ a, b กับสดมภ์ที่ c, d จะคำนวณอัตราส่วนออดส์ได้เท่ากับ

$$\frac{P_{ac}P_{bd}}{P_{bc}P_{ad}}$$

นอกจากนี้การคำนวณหาอัตราส่วนออดส์สำหรับตารางการณ์จรขนาด $r \times c$ สามารถพิจารณาจากแต่ละคู่ของแถว ซึ่งมีจำนวนทั้งหมด $\binom{r}{2} = \frac{r(r-1)}{2}$ คู่ และแต่ละคู่ของสดมภ์ ซึ่งมีจำนวนทั้งหมด $\binom{c}{2} = \frac{c(c-1)}{2}$ คู่ ดังนั้นเซตของอัตราส่วนออดส์ สามารถคำนวณได้จาก

$$\theta_{ij} = \frac{P_{i,j}P_{i+1,j+1}}{P_{i,j+1}P_{i+1,j}} \qquad ; \; i = 1,2,\ldots,r-1 \text{ และ } j = 1,2,\ldots,c-1$$

2.5 ฟังก์ชันภาวะน่าจะเป็นและตัวประมาณค่าภาวะน่าจะเป็นสูงสุด (Likelihood Function and Maximum Likelihood Estimator or MLE)

การเลือกตัวอย่างนั้นเป็นขั้นแรกในการวิเคราะห์ข้อมูล โดยการเลือกตัวอย่างนั้นเราไม่ทราบเซตของพารามิเตอร์ ในการวิเคราะห์ข้อมูลจึงต้องอาศัยการเลือกตัวแบบต่างๆ กัน เพื่อต้องการทราบถึงเซตของค่าพารามิเตอร์ที่สามารถอธิบายโครงสร้างความสัมพันธ์ของข้อมูลที่สนใจในการศึกษา กำหนดให้ค่าสังเกต $y = (y_1, y_2, \ldots, y_N)'$ เป็นค่าจริงที่สอดคล้องกับเซตของตัวแปรเชิงสุ่ม $Y = (Y_1, Y_2, \ldots, Y_N)'$ ที่มีฟังก์ชันความหนาแน่นของความน่าจะเป็นร่วม (Joint Probability Density Function) ที่มี $\theta_1, \theta_2, \ldots, \theta_p$ เป็นค่าพารามิเตอร์ คือ

$$f(Y;\theta) = f(U_1, Y_2, \ldots, Y_N; \theta_1, \theta_2, \ldots, \theta_p)$$

โดยที่ Y แทน $(Y_1, Y_2, \ldots, Y_N)'$ และ θ แทน $(\theta_1, \theta_2, \ldots, \theta_p)'$ และฟังก์ชันภาวะน่าจะเป็นของข้อมูลที่สังเกตได้คือ $\prod_{i=1}^{N} f(y_i|\theta)$

ในกรณีที่มีค่าสังเกตของตัวแปรเชิงสุ่ม Y จำนวน N ค่าที่เป็นอิสระต่อกัน กล่าวคือ $y_1, y_2, \ldots, y_N$ ภายใต้พารามิเตอร์ θ จะมีฟังก์ชันภาวะน่าจะเป็น คือ

$$f(y;\theta) = f(y_1, y_2, \ldots, y_N; \theta_1, \theta_2, \ldots, \theta_p)$$

$$= f(y_1;\theta_1)\, f(y_2;\theta_2) \ldots f(y_p;\theta_p)$$

กำหนดให้ $L = \ln f(y;\theta) = \sum_{i=1}^{N} f(y_i;\theta)$ และ Ω แทนเซตที่เป็นไปได้ทั้งหมดของเวกเตอร์พารามิเตอร์ θ

สำหรับตัวประมาณค่าภาวะน่าจะเป็นสูงสุด (Maximum Likelihood Estimator or MLE) คือค่าพารามิเตอร์ที่ทำให้ฟังก์ชันภาวะน่าจะเป็นมีค่าสูงสุด ซึ่งตัวประมาณค่าภาวะน่าจะเป็นสูงสุดของ θ คือ $\hat{\theta}$ ที่ทำให้ฟังก์ชันภาวะน่าจะเป็น $f(y;\theta)$ มีค่าสูงสุด นั่นคือ

$$f(y;\hat{\theta}) \geq f(y;\theta) \qquad \text{สำหรับทุกค่า } \theta \text{ ใน } \Omega$$

ค่าประมาณภาวะน่าจะเป็นสูงสุด $(\hat{\theta})$ เป็นค่าที่ทำให้ $L(\theta)$ มีค่าสูงที่สุดด้วย และฟังก์ชันลอการิทึมเป็นฟังก์ชันที่มีผลในทางเดียวกัน ดังนั้น

$$\ln f(y;\hat{\theta}) \geq \ln f(y;\theta) \qquad \text{สำหรับทุกค่า } \theta \text{ ใน } \Omega$$

หรือ $$L(\hat{\theta}) \geq L(\theta) \qquad \text{สำหรับทุกค่า } \theta \text{ ใน } \Omega$$

โดยทั่วไปนิยมใช้ฟังก์ชันของ Log-Likelihood มากกว่า ฟังก์ชันของ Likelihood โดยตรง เนื่องจากสามารถช่วยในการคำนวณให้สะดวกมากขึ้น สำหรับตัวประมาณภาวะน่าจะเป็นสูงสุด สามารถหาได้จากการหาอนุพันธ์ของฟังก์ชัน L เทียบกับพารามิเตอร์ทีละตัวและทำการเทียบให้เท่ากับศูนย์ แล้วทำการแก้สมการพร้อมกัน ดังนี้

$$\frac{\partial}{\partial \theta_j} L = 0 \qquad \text{เมื่อ } j = 1, 2, ..., p$$

การตรวจว่าฟังก์ชัน L มีค่าสูงสุดหรือไม่นั้น สามารถทำได้โดยการหาเมตริกของอนุพันธ์อันดับสอง (Second Derivatives) ของ $L\left(\frac{\partial^2}{\partial \theta_j \partial \theta_k} L\right)$ ว่ามีค่าเป็น Negative Definite ณ ค่า $\hat{\theta}$ หรือไม่ เช่นในกรณีที่มีพารามิเตอร์ θ เพียงตัวเดียว การตรวจสอบว่าฟังก์ชัน L มีค่าสูงสุดหรือไม่นั้น สามารถทำได้โดยการหาค่าของ $\frac{\partial^2}{\partial \theta^2} L$ ว่ามีค่าเป็นลบหรือไม่ เมื่อแทนค่าด้วย $\hat{\theta}$

2.6 การทดสอบภาวะรูปดีแบบต่างๆ

2.6.1 การทดสอบความเป็นอิสระโดยใช้การทดสอบไคสแควร์ (Chi-Square Test)

การทดสอบไคสแควร์เป็นการทดสอบสมมติฐาน ในกรณีที่ข้อมูลประกอบด้วยสองตัวแปรที่เป็นตัวแปรเชิงกลุ่ม ซึ่งข้อมูลอาจเป็นนามบัญญัติหรือเรียงลำดับก็ได้ และสามารถนำมาจัดลงในตารางการณ์จร สมมติว่าตัวแปรทั้งสองสามารถแบ่งออกได้เป็น r และ c กลุ่มตามลำดับ

กำหนดให้ O_{ij} เป็นจำนวนค่าสังเกตที่อยู่ในกลุ่มที่ i ของตัวแปร X และกลุ่มที่ j ของตัวแปร Y โดยที่ $i = 1, 2, ..., r$ และ $j = 1, 2, ..., c$ ดังตาราง 2.3

ตาราง 2.3 ค่าสังเกตที่อยู่ในกลุ่มที่ i ของตัวแปร X และกลุ่มที่ j ของตัวแปร Y

ตัวแปร X	ตัวแปร Y				รวม
	Y_1	Y_2	…	Y_c	
X_1	O_{11}	O_{12}	…	O_{1c}	$n_{1\cdot}$
X_2	O_{21}	O_{22}	…	O_{1c}	$n_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$	O_{ij}	$\vdots$	$\vdots$
X_r	O_{r1}	O_{r2}	…	O_{rc}	$n_{r\cdot}$
รวม	$n_{\cdot 1}$	$n_{\cdot 2}$	…	$n_{\cdot c}$	$n_{\cdot\cdot} = n$

มีสมมติฐานเพื่อการทดสอบความเป็นอิสระระหว่างตัวแปรเชิงกลุ่ม 2 ตัวแปร ดังนี้

H_0 : ตัวแปร X และ Y เป็นอิสระต่อกัน

H_1 : ตัวแปร X และ Y ไม่เป็นอิสระต่อกัน

จากทฤษฎีความน่าจะเป็นของ 2 ตัวแปรที่เป็นอิสระต่อกัน คือ

$$\begin{aligned} p_{ij} &= p_{i\cdot} p_{\cdot j} \\ &= \left(\frac{n_{i\cdot}}{n}\right)\left(\frac{n_{\cdot j}}{n}\right) \end{aligned}$$

ให้ E_{ij} เป็นค่าคาดหวังของกลุ่ม X_i และ Y_j โดยที่ $i = 1, 2, ..., r$ และ $j = 1, 2, ..., c$ ซึ่ง

$$E_{ij} = np_{ij} = n\left(\frac{n_{i\cdot}}{n}\right)\left(\frac{n_{\cdot j}}{n}\right) = \frac{n_{i\cdot} n_{\cdot j}}{n} = m_{ij}$$

ซึ่งตัวสถิติที่ใช้ในการทดสอบคือ

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}}$$

จะปฏิเสธสมมติฐาน H_0 ถ้าค่า χ^2 ที่ได้จากการคำนวณมีค่ามากกว่า $\chi^2_{\alpha,(i-1)(j-1)}$ จากตารางการแจกแจงแบบไคสแควร์ที่องศาความเป็นอิสระ (Degree of freedom: df) เท่ากับ (r–1)(c–1) ที่ระดับนัยสำคัญ (Significant level:α) ที่กำหนดไว้

ในกรณีที่ค่าคาดหวังในแต่ละช่องมีบางตัวมีค่าน้อยกว่า 5 ซึ่งมีจำนวนกลุ่มเกิน 20% ของจำนวนกลุ่มทั้งหมด ไม่ควรใช้สถิติไคสแควร์มาทดสอบโดยตรง เพราะค่าที่คำนวณได้จะมีค่าไม่แกร่งพอ (Robust) ซึ่งอาจแก้ไขโดยการยุบเพื่อรวมกลุ่มที่อยู่ติดกัน (Adjacent cells) เพื่อให้ค่าคาดหวังมีค่ามากกว่า 5 และสำหรับในกรณีที่ระดับองศาความเป็นเป็นอิสระเท่ากับ 1 หรือข้อมูลอยู่ในรูปของตารางการณ์จรขนาด 2×2 จะแก้ไขด้วยวิธีการของเยทส์ (Yates' Correction) ดังนี้

$$\chi^2 = \sum_{i=1}^{2}\sum_{j=1}^{2}\frac{\left(\left|O_{ij} - E_{ij}\right| - 0.5\right)^2}{E_{ij}}$$

จะปฏิเสธสมมติฐาน H_0 ถ้าค่า χ^2 ที่คำนวณได้มีค่ามากกว่า χ^2_1 จากตารางการแจกแจงแบบไคสแควร์ที่องศาความเป็นอิสระเท่ากับ 1 ที่ระดับนัยสำคัญที่กำหนดไว้

การทดสอบไคสแควร์ นอกจากจะใช้ทดสอบความสัมพันธ์หรือความเป็นอิสระระหว่างตัวแปรแล้ว ยังสามารถใช้ในการทดสอบการแจกแจงของประชากรว่าเป็นไปตามการแจกแจงที่คาดไว้หรือไม่ ซึ่งเรียกว่า การทดสอบความเหมาะสมของตัวแบบหรือการทดสอบภาวะสารูปสนิทดี (Goodness of fit)

2.6.2 ตัวสถิติอัตราส่วนภาวะน่าจะเป็นไคสแควร์ (Likelihood-Ratio Chi-Squared Statistic)

เป็นวิธีการที่สามารถใช้ทดสอบสมมติฐานหลักและสมมติฐานรองที่มีวัตถุประสงค์ต่างๆ กัน เช่น การทดสอบภาวะรูปดี การทดสอบความเป็นอิสระ การทดสอบพารามิเตอร์ว่าเป็นศูนย์หรือไม่ การเปรียบเทียบตัวแบบสองตัวแบบ เป็นต้น

สำหรับตัวสถิติอัตราส่วนภาวะน่าจะเป็นไคสแควร์นั้น Wilks ได้ทำการพิสูจน์ว่า $-2\ln\Lambda$ มีการแจกแจงแบบไคสแควร์ โดยกำหนดให้ Λ แทนอัตราส่วนภาวะน่าจะเป็นสูงสุด (Maximized Likelihood) ซึ่งจะมีค่าน้อยกว่าหรือเท่ากับ 1 และกรณีทั่วไปของตารางการณ์จรสองทาง จะมีรูปแบบของอัตราส่วนภาวะน่าจะเป็น ดังนี้

$$\Lambda = \frac{\prod_{i=1}^{r}\prod_{j=1}^{c}(n_{i\cdot}n_{\cdot j})^{n_{ij}}}{n^n\prod_{i=1}^{r}\prod_{j=1}^{c}n_{ij}^{n_{ij}}}$$

และตัวสถิติของ Wilks หรือตัวสถิติอัตราส่วนภาวะน่าจะเป็นไคสแควร์ คือ

$$G^2(M) = -2\ln\Lambda$$

$$\therefore G^2(M) = 2\sum_{i=1}^{r}\sum_{j=1}^{c} n_{ij}\ln\left(\frac{n_{ij}}{\hat{m}_{ij}}\right)$$

โดยที่ $\hat{m}_{ij} = \dfrac{n_{i\cdot}n_{\cdot j}}{n}$ เป็นตัวประมาณความถี่คาดหวังภายใต้ข้อสมมติเกี่ยวกับความเป็นอิสระ

เมื่อข้อมูลเชิงกลุ่มเป็นอิสระต่อกัน $G^2(M)$ จะมีการแจกแจงแบบ Asymptotic ไคสแควร์ด้วยองศาความเป็นอิสระเท่ากับ (r-1)(c-1)

นอกจากนี้ยังใช้หลักการของอัตราส่วนภาวะน่าจะเป็นสร้างตัวสถิติ $-2(L_2-L_1)$ ซึ่งใช้ในการทดสอบพารามิเตอร์ของตัวแบบว่ามีค่าเป็นศูนย์หรือไม่ นั่นคือ การเปรียบเทียบตัวแบบที่กำหนดขึ้นคือ ตัวแบบ M_1 กับตัวแบบที่มีจำนวนพารามิเตอร์น้อยกว่าตัวแบบ M_1 คือตัวแบบ M_2

ในการวิเคราะห์ข้อมูลเชิงกลุ่ม จะให้ $G^2(M_2|M_1)$ แทนตัวสถิติสำหรับการทดสอบตัวแบบ M_2 เมื่อกำหนดตัวแบบ M_1 ส่วนการทดสอบภาวะรูปดีของ $G^2(M)$ เป็นกรณีพิเศษของ $G^2(M_2|M_1)$ โดยที่ M_2 = M และ M_1 เป็นตัวแบบเต็มรูป (Saturated) ในการทดสอบตัวแบบ M จะทดสอบว่าพารามิเตอร์ทั้งหมดที่มีอยู่ในตัวแบบ M_1 แต่ไม่อยู่ในตัวแบบ M_2 ว่ามีค่าเป็นศูนย์หรือไม่ การทดสอบจะแสดงโดยใช้สัญลักษณ์เพิ่มเติมดังนี้

ให้ L_p แทน Maximized Likelihood ของตัวแบบเต็มรูป ดังนั้น ตัวสถิติอัตราส่วนภาวะน่าจะเป็น สำหรับการเปรียบเทียบตัวแบบ M_1 และตัวแบบ M_2 คือ

$$\begin{aligned} G^2(M_2|M_1) &= -2(L_2-L_1) \\ &= -2(L_2-L_p)-(-2)(L_1-L_p) \\ &= G^2(M_2)-G^2(M_1) \end{aligned}$$

2.7 ตัวแบบเชิงเส้นที่วางนัยทั่วไป (Generalized Linear Model or GLM)

GLM เป็นตัวแบบที่ขยายจากตัวแบบเชิงเส้นแบบคลาสสิค (Classical Linear Model) ซึ่งสามารถประยุกต์ใช้ได้ทั้งข้อมูลที่มีลักษณะต่อเนื่องและไม่ต่อเนื่อง โดยมีส่วนประกอบ 3 ส่วน คือ ส่วนประกอบเชิงสุ่ม (Random Component) ส่วนประกอบแบบมีระบบ (Systematic Component) และส่วนประกอบที่เชื่อมฟังก์ชันความสัมพันธ์ (Link Function) มีรายละเอียดดังนี้

1. ส่วนประกอบเชิงสุ่ม (Random Component) เป็นส่วนประกอบที่เกี่ยวข้องกับสมมติฐานของการแจกแจงของตัวแปรสุ่ม (Y) ที่เป็นตัวแปรตาม โดยตัวแปรสุ่มนอกจากจะมีการแจกแจงแบบปกติเหมือนในตัวแบบเชิงเส้นแบบคลาสสิคแล้วยังสามารถขยายไปสู่การแจกแจง

แบบอื่นซึ่งอยู่ในกลุ่มเอกซ์โปเนนเชียล (Exponential Family) เช่นการแจกแจกแบบปัวส์ซง การแจกแจงแบบทวินามและการแจกแจงแบบแกมมา เป็นต้น โดยสามารถแสดงฟังก์ชันความน่าจะเป็นของ $(y_1, y_2, ..., y_n)$ ได้ดังนี้

$$f(y_i : \theta_i, \phi) = \exp\left[\left(\frac{y_i\theta_i - b(\theta_i)}{a_i(\phi)}\right) + c(y_i, \phi)\right]$$

โดยที่ θ_i คือ Natural Parameter

$b(\theta_i), c(y_i, \phi)$ คือ ลักษณะเฉพาะของฟังก์ชันในกลุ่มเอกซ์โปเนนเชียล

ϕ คือ Dispersion Parameter อาจทราบค่าหรือไม่ทราบค่าก็ได้

ในกรณีที่ทราบค่า ϕ ตัวแบบที่ได้จะอยู่ในกลุ่มตัวแบบที่มีการแจกแจงเอกซ์โปเนนเชียลที่มีพารามิเตอร์ θ_i แต่ถ้าไม่ทราบค่า ϕ ตัวแบบที่ได้อาจเป็นหรือไม่เป็นไม่เป็นตัวแบบที่อยู่ในกลุ่มตัวแบบที่มีการแจกแจงเอกซ์โปเนนเชียล ที่มีพารามิเตอร์ 2 ตัว (θ_i, ϕ) โดยมีค่าเฉลี่ยเท่ากับ $b'(\theta_i)$ และความแปรปรวนเท่ากับ $b''(\theta_i)a_i(\phi)$

2. ส่วนประกอบแบบมีระบบ (Systematic Component) เป็นส่วนเชื่อมเวกเตอร์ η (Linear predictor) หรือตัวพยากรณ์เชิงเส้นกับเซตของตัวแปรอิสระ ให้มีรูปแบบเชิงเส้น ดังนี้

$$\underset{\sim}{\eta} = X\underset{\sim}{\beta} ; \eta_i = \sum_{i=1}^{N} \beta_j X_{ij} \qquad ,i = 1,2,...,N \quad j = 1,2,...,p$$

โดยที่ X แทนเมตริกซ์ของตัวแปรอิสระที่ประกอบด้วยค่าสังเกตขนาด N

$\underset{\sim}{\beta}$ แทนเวกเตอร์ของพารามิเตอร์ $(\beta_1, ..., \beta_p)$

$\underset{\sim}{\eta}$ แทนเวกเตอร์ของตัวพยากรณ์เชิงเส้น

p แทนจำนวนของตัวแปรอิสระ

3. ส่วนประกอบที่เชื่อมฟังก์ชันความสัมพันธ์ (Link Function) สำหรับเชื่อมส่วนประกอบเชิงสุ่มและส่วนประกอบแบบมีระบบเข้าด้วยกัน เพื่ออธิบายความสัมพันธ์ระหว่างตัวแปรตามกับตัวแปรอิสระ นอกจากจะใช้ Identity Link หรือเรียกอีกอย่างว่า Canonical Link $\left(\underset{\sim}{\eta} = \mu_i\right)$ ซึ่งอยู่ในตัวแบบเชิงเส้นแบบคลาสสิคแล้ว ยังสามารถใช้เชื่อมความสัมพันธ์กับฟังก์ชันอื่นๆ อีกหลายแบบที่เป็นฟังก์ชันแบบ Monotonic differentiable function เช่น Log Link $\left(\underset{\sim}{\eta} = \log(\mu_i)\right)$ โดยการใช้ Link Function จะขึ้นอยู่กับการแจกแจงของตัวแปรตาม นอกจากนี้ยังพบว่า Link Function ของแต่ละการแจกแจงคือ θ_i

2.8 งานวิจัยที่เกี่ยวข้อง

กัลยาณี ตันตรานนท์ (2547) ทำการศึกษาการสูญเสียการได้ยินของคนงานและการใช้อุปกรณ์ป้องกันเสียง ในโรงงานผลิตอาหารกระป๋องขนาดใหญ่ จังหวัดเชียงใหม่ พบว่า อัตราความชุกของการสูญเสียการได้ยินของคนงานในโรงงานที่ศึกษาเท่ากับร้อยละ 21.0 และ การสูญเสียการได้ยินไม่มีความสัมพันธ์กับความถี่ของการสวมใส่อุปกรณ์ป้องกันเสียง แต่มีความสัมพันธ์กับความถูกต้องของวิธีการใช้อุปกรณ์ป้องกันเสียง อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05

ทวีพร บุญวานิช (2541) ทำการศึกษาการประยุกต์โมเดลล็อกลิเนียร์ในการวิเคราะห์สาเหตุเพื่อศึกษาปัจจัยที่ส่งผลต่อระยะเวลาที่ใช้ในการศึกษาของมหาบัณฑิตทางสังคมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย มีวัตถุประสงค์เพื่อประยุกต์โมเดลล็อกลิเนียร์ในการวิเคราะห์สาเหตุของอิทธิพลหลักและปฏิสัมพันธ์อันดับหนึ่งของปัจจัยที่ส่งผลทางตรง ทางอ้อม (ได้แก่ อายุ ลักษณะการลาศึกษา คุณลักษณะอาจารย์ที่ปรึกษาวิทยานิพนธ์ การบริการของหน่วยงาน สมรรถภาพการทำวิทยานิพนธ์ และปัญหาการทำวิทยานิพนธ์) ต่อระยะเวลาที่ใช้ในการศึกษาของมหาบัณฑิตทางสังคมศาสตร์ กลุ่มตัวอย่างในการวิจัยได้แก่ มหาบัณฑิตทางสังคมศาสตร์ รุ่นปีการศึกษา 2535 จำนวน 346 คน และรุ่นปีการศึกษา 2536 จำนวน 403 คน พบว่า ในนิสิตรุ่น 2535 ปัจจัยที่ส่งผลต่อระยะเวลาที่ใช้ในการศึกษาอย่างมีนัยสำคัญทางสถิติ มีเฉพาะอิทธิพลทางตรงเท่านั้น ได้แก่ อิทธิพลหลักของลักษณะการลาศึกษาและปฏิสัมพันธ์อันดับหนึ่งระหว่างอายุกับการบริการของหน่วยงาน ระหว่างลักษณะการลาศึกษากับปัญหาการทำวิทยานิพนธ์และระหว่างคุณลักษณะอาจารย์ที่ปรึกษาวิทยานิพนธ์กับปัญหาการทำวิทยานิพนธ์ ส่วนนิสิตรุ่น 2536 มีทั้งทางตรงและทางอ้อม ได้แก่ อิทธิพลหลักของลักษณะการลาศึกษา คุณลักษณะอาจารย์ที่ปรึกษาวิทยานิพนธ์และการบริการของหน่วยงาน และปฏิสัมพันธ์อันดับหนึ่งระหว่างการบริการของหน่วยงานกับปัญหาการทำวิทยานิพนธ์ ระหว่างอายุกับลักษณะการลาศึกษา และระหว่างอายุกับคุณลักษณะของอาจารย์ที่ปรึกษาวิทยานิพนธ์

นิรมล นราวิวัฒน์ (2541) ทำการศึกษาการสูญเสียการได้ยินในคนงานโรงงานอุตสาหกรรมผลิตกระดาษ ในกรุงเทพมหานคร มีวัตถุประสงค์เพื่อจะทราบความชุกของการสูญเสียสมรรถภาพการได้ยิน และศึกษาสัมพันธ์ของการสูญเสียการได้ยินกับระดับความดังของเสียงอายุ อายุการทำงาน ระดับการศึกษา และแผนกงาน โดยสุ่มตัวอย่างจำนวน 165 คน ที่ทำงานในแผนกที่มีเสียงดังเกิน 85 dB(A) ขึ้นไป พบว่า ระดับความดังของเสียงทุกแผนกมีเสียงดังเกิน 85 dB(A) ยกเว้นห้องควบคุม (DCS Control Room) ชั้นล่างมีเสียงดัง 73.4 dB(A) และคนงานมีอัตราการสูญเสียการได้ยินร้อยละ 25.6 พร้อมการวิเคราะห์ระหว่างปัจจัยต่างๆ กับการสูญเสียการได้ยินพบว่า การสูญเสียการได้ยินมีความสัมพันธ์กับอายุ อายุงาน และการศึกษา อย่างมีนัยสำคัญทางสถิติ

เบ็ญจมาศ แสงอนุเคราะห์ (2541) ทำการศึกษาการพัฒนาโมเดลความคาดหวังในการศึกษาต่อของนักเรียนชั้นมัธยมศึกษาปีที่ 3 โรงเรียนขยายโอกาสทางการศึกษาขั้นพื้นฐาน เขตการศึกษา 6:

การวิเคราะห์ล็อกลิเนียร์มาตรอันดับ มีวัตถุประสงค์เพื่อศึกษาความสัมพันธ์ของปัจจัยที่ส่งผลต่อความคาดหวังในการศึกษาต่อของนักเรียน เพื่อศึกษาอิทธิพลหลักและอิทธิพลปฏิสัมพันธ์ของปัจจัยที่ส่งผลต่อความคาดหวังในการศึกษาต่อของนักเรียนและเพื่อพัฒนาโมเดลล็อกลิเนียร์สำหรับตัวแปรมาตรอันดับของความคาดหวังในการศึกษาต่อของนักเรียน โดยกลุ่มตัวอย่างคือนักเรียนชั้นมัธยมศึกษาปีที่ 3 ปีการศึกษา 2541 จำนวน 500 คน สุ่มเลือกมาจากประชากรในเขตการศึกษาเขต 6 ผลการศึกษาพบว่า ปัจจัยที่มีความสัมพันธ์กับความคาดหวังในการศึกษาต่อของนักเรียนชั้นมัธยมศึกษาปีที่ 3 โรงเรียนขยายโอกาสทางการศึกษาขั้นพื้นฐาน อย่างมีนัยสำคัญได้แก่ ผลสัมฤทธิ์ทางการเรียนของนักเรียน อาชีพของผู้ปกครอง รายได้ของผู้ปกครองและการศึกษาของผู้ปกครองตามลำดับ และปัจจัยที่ส่งผลต่อความคาดหวังในการศึกษาต่อของนักเรียนชั้นมัธยมศึกษาปีที่ 3 โรงเรียนขยายโอกาสทางการศึกษาขั้นพื้นฐาน อย่างมีนัยสำคัญ มีเฉพาะอิทธิพลหลักเท่านั้น ซึ่งได้แก่ อิทธิหลักของผลสัมฤทธิ์ทางการเรียนของนักเรียน การศึกษาของผู้ปกครองและอาชีพของผู้ปกครอง ตามลำดับ

พรพิศ ยิ้มประยูร (2545) ทำการศึกษาการประยุกต์ตัวแบบล็อก ลิเนียร์สำหรับผู้ป่วยโรคมะเร็ง: กรณีศึกษาสถาบันมะเร็งแห่งชาติ มีวัตถุประสงค์เพื่อหาความสัมพันธ์ระหว่างตัวแปรส่วนบุคคล ตัวแปรมะเร็งและตัวแปรเชิงคลินิก โดยกลุ่มตัวอย่างคือผู้ป่วยใหม่ที่เป็นโรคมะเร็งซึ่งเข้ารับการรักษาที่สถาบันมะเร็งแห่งชาติระหว่างเดือนมกราคม – ธันวาคม 2542 ผลการศึกษาพบว่าตัวแปรส่วนบุคคล ตัวแปรมะเร็งและตัวแปรเชิงคลินิกส่วนมากมีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติ โดยเฉพาะอย่างยิ่งตัวแปรตำแหน่งของการเป็นมะเร็งมีความสัมพันธ์กับตัวแปรสถานภาพการสมรส อายุ วิธีวินิจฉัยและการรักษา นอกจากนี้ตำแหน่งของการเป็นมะเร็งยังมีผลต่อวิธีวินิจฉัยและการรักษาด้วย ยิ่งไปกว่านั้น การรักษายังขึ้นอยู่กับอายุ ตำแหน่งของมะเร็งและระยะของการเป็นมะเร็งทั้งในผู้ป่วยหญิงและชาย

วาสนา เตื้องวิวัฒน์ (2542) ทำการศึกษาการประยุกต์ตัวแบบล็อก ลิเนียร์ สำหรับอุบัติเหตุการจราจรทางบกที่เกิดขึ้นในกรุงเทพมหานคร มีวัตถุประสงค์เพื่อหาความสัมพันธ์ระหว่างตัวแปรของการเกิดอุบัติเหตุการจราจรในแต่ละพื้นที่ โดยเลือกตัวอย่างจากสถานีตำรวจที่มีจำนวนอุบัติเหตุเกิดขึ้นบนท้องถนนมากที่สุดในแต่ละกองบังคับการเป็นตัวแทนในการศึกษาอุบัติเหตุจราจรทางบกที่เกิดขึ้น ในกรุงเทพมหานคร ปี พ.ศ. 2539 ซึ่งได้แก่ สถานีตำรวจนครบาลปทุมวัน ราษฎร์บูรณะและบางเขน จากกองบังคับการตำรวจนครบาลพระนครใต้ ธนบุรีและพระนครเหนือ ตามลำดับ ผลการศึกษาพบว่า ในแต่ละพื้นที่ ตัวแปรส่วนใหญ่มีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติ โดยเฉพาะอย่างยิ่งตัวแปร สาเหตุจากบุคคลที่ทำให้เกิดอุบัติเหตุ เช่น ขับรถเร็วเกินกว่าที่กฎหมายกำหนด ขับรถตามกระชั้นชิดและขับรถตัดหน้ากระชั้นชิด

สุทธิเดช ตัณฑ์พาณิชย์ (2543) ทำการศึกษาการวิเคราะห์ข้อมูลเชิงกลุ่ม พบว่า ตัวแบบล็อกลิเนียร์ เป็นตัวแบบที่มีตัวแปรอิสระเป็นตัวแปรเชิงกลุ่มและตัวแปรตามเป็นจำนวนนับในแต่ละช่องในตารางหลายทาง ตัวแบบดังกล่าวสามารถวิเคราะห์ความสัมพันธ์ของตัวแปรครั้งละหลายตัวแปร และสามารถอธิบายลักษณะความสัมพันธ์ระหว่างตัวแปรเชิงกลุ่มได้

Green JA. (1998) ทำการศึกษาการวิเคราะห์ล็อกลิเนียร์ของข้อมูลในตารางการณ์จรที่ตัวแปรมีมาตรวัดแบบเรียงลำดับ (Cross-classified ordinal data) พบว่า ตัวแบบล็อกลิเนียร์มีประโยชน์มากกว่าการทดสอบความเป็นอิสระของไคสแควร์ นั่นคือสามารถวิเคราะห์ปฏิสัมพันธ์ได้มากกว่า 2 ตัวแปร วัดความสัมพันธ์ระหว่างตัวอย่างทั้งตัวแบบเชิงเส้นหรือตัวแบบไม่เชิงเส้น และสามารถแปลผลคะแนนให้เป็นตัวแปรลำดับ (Ordinal Variable)