



## การใช้ตัวแบบเชิงเส้นน้อยทั่วไปในการวิเคราะห์ปัจจัยเสี่ยงต่อการเสียชีวิต ของทารกที่พิจารณาจากน้ำหนักแรกเกิดของทารก

### Using Generalized Linear Models in Analysis of Risk Factors to Infant Death Considering from Birth Weight of Infant

จุฑามาศ สลิโรจน์<sup>1</sup> ขวัญรัตน์ ขิตธิกุล<sup>1</sup> ดารารัตน์ ศรีเดช<sup>1</sup> และ วีรานันท์ พงศาภักดี<sup>1\*</sup>

#### บทคัดย่อ

การวิเคราะห์ปัจจัยที่เสี่ยงต่อการเสียชีวิตของทารกซึ่งพิจารณาจากน้ำหนักแรกเกิดของทารกจำนวน 3 กลุ่มคือ กลุ่มน้ำหนักทารกแรกเกิดที่เสี่ยงต่อการเสียชีวิต กลุ่มที่ต้องเฝ้าระวัง และกลุ่มปกติ โดยศึกษาปัจจัยเสี่ยงดังกล่าวจากวรรณกรรมที่เกี่ยวข้องได้แก่ อายุของมารดาที่ตั้งครรภ์ การสูบบุหรี่ของมารดาขณะตั้งครรภ์ การคลอดก่อนกำหนด ประวัติโรคความดันโลหิตสูงของมารดา การมีปัญหาสูตินารีเวช โดยใช้ข้อมูลจริง 189 คนจาก Hosmer and Lemeshow (1989) ด้วยตัวแบบลอจิสติก ตัวแบบลอจิสติก 2 กลุ่ม ตัวแบบล็อกลิเนียร์ ตัวแบบการถดถอยปัวซอง และตัวแบบการถดถอยปัวซองน้อยทั่วไป ผลการวิจัยพบว่า ทุกตัวแบบของการทดลองมีภาวะสารูปตัวอย่างมีนัยสำคัญทางสถิติที่ 0.05 ปัจจัยเสี่ยงต่อการเสียชีวิตของทารกแรกเกิดที่พบจากตัวแบบลอจิสติกและตัวแบบลอจิสติก 2 กลุ่มนั้นตรงกัน 3 ปัจจัยคือ การคลอดก่อนกำหนด ประวัติโรคความดันโลหิตสูงของมารดา และการมีปัญหาสูตินารีเวช ปัจจัยเสี่ยงจากตัวแบบล็อกลิเนียร์คือ ประวัติการเป็นโรคความดันโลหิตสูงของมารดา และการมีปัญหาสูตินารีเวช ส่วนปัจจัยเสี่ยงจากตัวแบบการถดถอยปัวซองเปรียบเทียบกับตัวแบบการถดถอยปัวซองน้อยทั่วไปคือ อายุของมารดาที่ตั้งครรภ์ ดังนั้นเกือบทุกปัจจัยเสี่ยงที่ศึกษาเกี่ยวกับการสูบบุหรี่ของมารดาขณะตั้งครรภ์ส่งผลกระทบต่อน้ำหนักแรกเกิดอย่างมีนัยสำคัญทางสถิติที่ 0.05 ด้วยค่าพยากรณ์ความน่าจะเป็นจากตัวแบบที่สอดคล้องกันค่อนข้างสูง

<sup>1</sup>ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยศิลปากร วิทยาเขตพระราชวังสนามจันทร์ นครปฐม 73000

\*Corresponding Author, E-mail: veeranun@su.ac.th

## ABSTRACT

The analysis of risk factors to infant death considering from birth weight of infant has been performed by classifying the birth weight into 3 categories: risky group, monitored group, and normal group. The risk factors obtained from literature reviews consist of mother age, smoking during pregnancy, premature labor, historical hypertension, and uterine irritability from the real data set of 189 babies (Hosmer and Lemeshow, 1989). The models under consideration are cumulative logit model, logit model for binary response, loglinear model, and Poisson regression model compared with generalized Poisson regression model. The results reveal that every model provides significance at 0.05 for goodness-of-fit test. The three risk factors affecting the response of cumulative logit model and the logit model are the same: premature labor, historical hypertension and uterine irritability. From loglinear model, history of hypertension and presence of uterine irritability are obtained. From Poisson regression model, the age of mothers is obtained. Hence almost all risk factors under studies, except for the smoking during pregnancy, are statistically affecting the birth-weight risk of death significantly at 0.05 with a highly predicted probability of a corresponding model.

**คำสำคัญ:** ทารกแรกเกิด ตัวแบบลอจิสต์ 2 กลุ่ม ตัวแบบลอจิสต์สะสม ตัวแบบล็อกลิเนียร์ ตัวแบบการถดถอยปัวซองทั่วไป

**Keyword:** Infant, Cumulative logit model, Logit model for binary response, Loglinear model, Generalized Poisson regression model

## บทนำ (introduction)

ทารกแรกเกิดหมายถึงทารกที่คลอดออกจากครรภ์มารดาตั้งแต่ 1 วันถึง 4 สัปดาห์ ทารกแรกเกิดเป็นวัยที่มีความสำคัญมากของมนุษย์ที่เริ่มเรียนรู้ที่จะเข้าใจหรือไว้วางใจในสภาพแวดล้อมรอบตัว และเป็นช่วงการเจริญเติบโตอย่างรวดเร็ว ปัจจุบันความรู้ทางการแพทย์เจริญขึ้นมาก จึงมีการจัดทารกที่คลอดตามเวลาตั้งครรภ์และน้ำหนักตัวเมื่อคลอดเป็น 5 ประเภทได้แก่ 1) ทารกปกติมีอายุในครรภ์มารดา 38-42 สัปดาห์ โดยน้ำหนักตัวของเด็กชายจะอยู่ในเกณฑ์ 3,100 กรัม ส่วนของเด็กหญิงอยู่ในเกณฑ์ 3,000 กรัม 2) ทารกที่ไม่ปกติที่มีอายุหรืออยู่ในครรภ์มารดาต่ำกว่า 3 สัปดาห์ เรียกว่าคลอดก่อนกำหนด 3) ทารกอยู่ในครรภ์มารดาเกิน 42 สัปดาห์ เรียกว่าคลอดหลังกำหนด 4) ทารกที่มีน้ำหนักน้อย ไม่ว่าจะอยู่ในครรภ์มารดานานเท่าใดก็ตาม เมื่อคลอดออกมาแล้วมีน้ำหนักตัวต่ำกว่า 2,500 กรัม เรียกว่าทารกมีน้ำหนักน้อย 5) ทารกที่น้ำหนักตัวน้อยกว่าเกณฑ์ของอายุครรภ์มารดาจะมีน้ำหนักต่ำกว่าร้อยละ 4 ของทารกโดยเฉลี่ย (<http://guru.sanook.com> สืบค้น มกราคม 2556) นอกจากนี้การประเมินผลการเติบโตของทารกแรกเกิด จะขึ้นอยู่กับน้ำหนักของทารกซึ่งมีความสำคัญอย่างมาก พ่อแม่จึงต้องติดตามและสนใจอย่างต่อเนื่อง เพื่อจะได้รู้ว่าลูกมีการเจริญเติบโตที่ปกติหรือไม่ โดยปกติทารกแรกเกิดมี

ค่าเฉลี่ยน้ำหนักตัวอยู่ที่ 2,800-3,200 กรัม หากน้ำหนักตัวทารกแรกเกิดนั้น น้อยกว่า 2,500 กรัม ถือว่าทารกนั้นมีน้ำหนักตัวต่ำกว่าเกณฑ์มาตรฐาน (<http://www.anamai.moph.go.th/main.php?filename=index> สืบค้น มกราคม 2556) ซึ่งอาจเกิดขึ้นกับทารกที่ครบกำหนดหรือคลอดก่อนกำหนดได้ ทารกคลอดน้ำหนักตัวน้อยส่วนใหญ่อาจเกิดมาจากสาเหตุเช่น มารดามีอายุน้อยกว่า 18 ปี หรือมากกว่า 35 ปี โรคที่มารดาเป็นอยู่เดิม เช่น โรคเลือด โรคความดันโลหิตสูง รวมถึงภาวะแทรกซ้อนระหว่างการตั้งครรภ์ เช่น ครรภ์เป็นพิษ และพฤติกรรมของมารดา เช่น การสูบบุหรี่ การดื่มสุรา และความผิดปกติของมดลูก จากรายงานการศึกษาทารกซึ่งเจริญเติบโตช้าในครรภ์ ที่โรงพยาบาลราชวิถีระหว่างเดือนมิถุนายน พ.ศ. 2527 ถึงเดือนพฤษภาคม พ.ศ. 2528 ระยะเวลา 1 ปี มีทารกน้ำหนักตัวน้อยเป็นร้อยละ 9.15 ของทารกคลอดมีชีวิตทั้งหมด และมีทารกเจริญเติบโตช้าในครรภ์ร้อยละ 19.35 ของทารกน้ำหนักตัวน้อย ซึ่งพบว่ามารดาของทารกที่เติบโตช้าในครรภ์นั้นมีประวัติการสูบบุหรี่ ภาวะของมดลูกไม่ปกติ ภาวะครรภ์เป็นพิษ การคลอดก่อนกำหนด และภาวะอื่น ๆ

งานวิจัยอื่น ๆ ซึ่งศึกษาปัจจัยที่มีผลต่อน้ำหนักทารกแรกเกิด เช่น งานผู้ป่วยในเด็กแรกคลอด โรงพยาบาลพนมสารคาม (กรมอนามัย กระทรวงสาธารณสุข) จังหวัดฉะเชิงเทรา รายงานปัจจัยที่มีความสัมพันธ์ต่อน้ำหนักทารกที่มารับบริการคือ อายุ อาชีพ ระดับการศึกษา จำนวนครั้งของการตั้งครรภ์ จำนวนการคลอด อายุครรภ์ ความสูงของยอดมดลูก น้ำหนักและส่วนสูงของมารดา จากตัวอย่างผู้มารับบริการในช่วงเดือนมิถุนายน-กันยายน พ.ศ. 2553 จำนวน 341 ราย ซึ่งมีพยาบาลเป็นผู้เก็บรวบรวมข้อมูล สำหรับสถิติที่ใช้วิเคราะห์ข้อมูล คือ สัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน (correlation of Pearson) ซึ่งแสดงให้เห็นว่า อายุครรภ์ น้ำหนัก ส่วนสูง และความสูงของยอดมดลูกมีความสัมพันธ์ต่อน้ำหนักทารกแรกเกิดอย่างมีนัยสำคัญ ที่ระดับ 0.05 ส่วนงานวิจัยในทำนองเดียวกันนี้เช่น งานของบุญสนอง (2550) ศึกษาปัจจัยที่มีความสัมพันธ์กับการเกิดทารกแรกเกิดน้ำหนักน้อยกว่า 2,500 กรัม ในเขต 9 เป็นการศึกษาแบบ retrospective study โดยใช้สถิติพรรณนา เช่น ร้อยละ และใช้สถิติวิเคราะห์ เช่น ตัวแบบการถดถอยพหุคูณ เพื่อหาความสัมพันธ์ระหว่างปัจจัยเสี่ยงด้านชีววิทยา ฐานะและสังคม อุปนิสัยในการดำเนินชีวิต ประวัติการตั้งครรภ์และการคลอดในอดีตและปัจจุบัน กับการเกิดทารกแรกเกิดน้ำหนักน้อยกว่า 2,500 กรัม และการเปรียบเทียบปัจจัยเสี่ยงระหว่างทารกแรกเกิดน้ำหนักน้อยกว่า 2,500 กรัม กับทารกแรกเกิดน้ำหนักปกติ ข้อมูลตัวอย่างคือมารดาครรภ์เดียวที่คลอดที่โรงพยาบาลศูนย์ และโรงพยาบาลทั่วไปทุกแห่งใน 6 จังหวัดของเขต 9 จำนวน 780 ราย ระหว่างวันที่ 1 กรกฎาคม – 31 สิงหาคม 2546 โดยใช้แบบบันทึกข้อมูลตามแบบรายงาน รบ.1 ต 05 แฟ้มรายงานการคลอดบุตร และการประเมินอายุครรภ์ของทารก ที่มีน้ำหนักแรกเกิดน้อยกว่า 2,500 กรัม ด้วย Ballard Score พบว่า ปัจจัยที่มีความสัมพันธ์กับน้ำหนักแรกเกิดน้อยกว่า 2,500 กรัมอย่างมีนัยสำคัญที่ระดับ 0.001 ได้แก่ อายุ ความดันโลหิตซิสโตลิกสูงกว่า 140 มิลลิเมตรปรอท การมีน้ำเดินก่อนกำหนด และภาวะของมดลูกผิดปกติ

ปัจจุบันยังมีปัญหาทารกที่มีน้ำหนักน้อยกว่าเกณฑ์อยู่มาก และยังไม่สามารถแก้ไขให้หมดไปได้ หากทารกยังมีน้ำหนักน้อยผิดปกติ อาจเกิดโรคแทรกซ้อนอื่น ๆ ด้วย ซึ่งอาจเป็นผลมาจากปัจจัยเสี่ยงต่าง ๆ มากมาย โดยส่งผลกระทบต่อทารกมีน้ำหนักน้อยผิดปกติ และอาจเสี่ยงต่อการเสียชีวิตของทารกได้ ผู้วิจัยจึงสนใจการวิเคราะห์ปัจจัยเสี่ยงที่มีผลต่อระดับน้ำหนักของทารกแรกเกิดที่เสี่ยงต่อการเสียชีวิต โดยอาศัยการวิเคราะห์ข้อมูลจำแนกประเภท โดยใช้ ตัวแบบเชิงเส้นนัยทั่วไป (generalized linear models: GLMs) (Agresti, 2002; วิ

รำนันท์, 2555) ที่สอดคล้องกับข้อมูลได้แก่ ตัวแบบลอจิสติกสะสม (cumulative logit model) (Walker and Duncan, 1967) ตัวแบบลอจิสติก 2 กลุ่ม (logit model for binary response) (Agresti, 2002) ตัวแบบล็อกลิเนียร์ (log-linear model) (Birch, 1963; Agresti, 2002) ตัวแบบการถดถอยปัวซอง (Poisson regression model) และตัวแบบการถดถอยปัวซองนัยทั่วไป (generalized Poisson regression model) (Agresti, 2002; เมธิยา, 2554) ด้วยการทดสอบภาวะสารถูปดีของตัวแบบทุกตัวแบบกับชุดข้อมูลจริงจาก Hosmer and Lemeshow (1989) เพื่อใช้ร่วมกับการหาสารสนเทศและปัจจัยเสี่ยงที่มีผลต่อกลุ่มน้ำหนักของทารกแรกเกิดที่เสี่ยงต่อการเสียชีวิต โดยเน้นศึกษาปัจจัยเสี่ยงต่าง ๆ ที่พบจากวรรณกรรมที่เกี่ยวข้องดังนี้ น้ำหนักของทารกแรกเกิด (birth weight in gram: Bwt) อายุมารดาที่ตั้งครรภ์ (mother age in years: Age) การสูบบุหรี่ของมารดาขณะตั้งครรภ์ (smoking during pregnancy: Smoke) การคลอดก่อนกำหนด (premature labor: Ptl) ประวัติโรคความดันโลหิตสูงของมารดา (historical hypertension: Ht) และ การมีปัญหาสุตินารีเวช (presence of uterine irritability: Ui) รวมทั้งการพยากรณ์ความน่าจะเป็นของกลุ่มที่เสี่ยงต่อการเสียชีวิตด้วยตัวแบบเชิงสถิติต่าง ๆ ที่เหมาะสม

### วิธีการดำเนินการวิจัย (research methodology)

ข้อมูล ตัวแปร และตัวแบบเชิงเส้นนัยทั่วไปที่ใช้ในการวิจัย ประกอบด้วยตัวแปรตอบสนอง และตัวแปรอธิบายหรือปัจจัยเสี่ยงเชิงกลุ่มต่าง ๆ ดังนี้

1. อายุมารดาที่ตั้งครรภ์ (mother age in years: Age) จำแนกเป็น 3 กลุ่มคือ อายุ 14-19 ปี (ให้เป็นกลุ่ม 0) อายุ 20-40 ปี (ให้เป็นกลุ่ม 1) อายุมากกว่า 41-60 ปี (ให้เป็นกลุ่ม 2)
2. การสูบบุหรี่ของมารดาขณะตั้งครรภ์ (smoking during pregnancy: Smoke) จำแนกเป็น 2 กลุ่มคือ ไม่สูบบุหรี่ (ให้เป็นกลุ่ม 0) สูบบุหรี่ (ให้เป็นกลุ่ม 1)
3. การคลอดก่อนกำหนด (premature labor: Ptl) จำแนกเป็น 3 กลุ่มคือ ไม่คลอดก่อนกำหนด (ให้เป็นกลุ่ม 0) คลอดก่อนกำหนดภายใน 1 เดือน (ให้เป็นกลุ่ม 1) คลอดก่อนกำหนดมากกว่า 1 เดือนขึ้นไป (ให้เป็นกลุ่ม 2)
4. ประวัติโรคความดันโลหิตสูงของมารดา (historical hypertension: Ht) จำแนกเป็น 2 กลุ่มคือ ไม่เป็นความดันโลหิตสูง (ให้เป็นกลุ่ม 0) เป็นความดันโลหิตสูง (ให้เป็นกลุ่ม 1)
5. การมีปัญหาสุตินารีเวช (presence of uterine irritability: Ui) จำแนกเป็น 2 กลุ่มคือ ไม่มีการกระทบกระเทือนมดลูก (ให้เป็นกลุ่ม 0) มีการกระทบกระเทือนมดลูก (ให้เป็นกลุ่ม 1)
6. น้ำหนักของทารกแรกเกิด (birth weight: Bwt) แทนตัวแปรตอบสนอง จำแนกเป็น 3 ระดับคือ
  - 1) น้ำหนักของเด็กแรกเกิดน้อยกว่าหรือเท่ากับ 2,001 กรัม อยู่ในระดับเสี่ยงต่อการเสียชีวิต (ให้เป็นกลุ่ม 1)
  - 2) น้ำหนักของเด็กแรกเกิด 2,000–2,499 กรัม อยู่ในระดับที่ต้องเฝ้าระวัง (ให้เป็นกลุ่ม 2)
  - 3) น้ำหนักของทารกแรกเกิดมากกว่าหรือเท่ากับ 2,500 กรัม อยู่ในระดับปกติ (ให้เป็นกลุ่ม 3)

ตัวแบบที่ใช้ในการวิจัย อาศัยตัวแบบเชิงเส้นน้อยทั่วไป โดยศึกษาตัวแบบต่าง ๆ และทฤษฎีที่เกี่ยวข้อง ซึ่งโดยทั่วไปตัวแบบ GLMs จะนิยมใช้สัญลักษณ์ log ที่หมายถึงกรณีเดียวกันกับ natural log ด้วย โดยแบ่งออกเป็น 4 กรณี ดังนี้

กรณีที่ 1: การวิเคราะห์ข้อมูลด้วยตัวแบบลอจิสติกสะสม (cumulative logit model) เมื่อพิจารณาตัวแปรตอบสนองเป็นแบบอันดับ (ordinal response variable) คือ Bwt แบ่งเป็น 3 กลุ่ม ดังข้างต้น

กรณีที่ 2: การวิเคราะห์ข้อมูลด้วยตัวแบบลอจิสติก 2 กลุ่ม (logit model for binary response) เมื่อตัวแปรตอบสนองแบ่งเป็น 2 กลุ่มคือ กลุ่มน้ำหนักของทารกแรกเกิด ( $\leq 2,499$  กรัม) แทนกลุ่มที่อยู่ในระดับเสี่ยงต่อการเสียชีวิตที่ต้องเฝ้าระวัง และกลุ่มน้ำหนักของทารกแรกเกิด ( $\geq 2,500$  กรัม) แทนกลุ่มที่อยู่ในระดับปกติ

กรณีที่ 3: การวิเคราะห์ข้อมูลโดยใช้ตัวแบบลอกลินีเยอร์ เมื่อตัวแปรตอบสนองแทนด้วยจำนวนนับภายใต้การแจกแจงปัวซอง

กรณีที่ 4: ตัวแบบการถดถอยปัวซองน้อยทั่วไป เมื่อพิจารณาตัวแปรตอบสนองแทนด้วยจำนวนนับภายใต้การแจกแจงทวินามลบ (negative binomial distribution) และเปรียบเทียบกับตัวแบบการถดถอยปัวซอง (Poisson regression model) ภายใต้การแจกแจงปัวซอง

ตัวแบบเชิงเส้นน้อยทั่วไป (generalized linear models หรือ GLMs) มีหลายตัวแบบเป็นกลุ่มของตัวแบบที่ขึ้นอยู่กับลักษณะการแจกแจงของตัวแปรตอบสนองที่อยู่ในวงศ์เลขชี้กำลัง (exponential family) ส่วนประกอบเชิงระบบ (systematic component) แบบเชิงเส้นในเทอมพารามิเตอร์ และฟังก์ชันเชื่อมโยง (link function) แบบต่าง ๆ รวมถึงจำนวนของตัวแปรอธิบายแบบต่าง ๆ เช่น การแจกแจงตัวแปรตอบสนองอันดับแบบพหุนาม (multinomial) ที่นิยมใช้ฟังก์ชันลอจิสติกสะสม (cumulative logit function) ในกรณี proportional odds model ซึ่งเสนอโดย McCullagh (1980) เป็นฟังก์ชันเชื่อมโยง ในงานวิจัยนี้มีตัวแบบ GLMs ที่เกี่ยวข้องดังนี้

ตัวแบบลอจิสติกสะสม เสนอโดย Walker and Duncan (1967) ซึ่งต่อมาได้ถูกปรับใช้ในกรณีพิเศษที่เรียกว่า proportional odds model (McCullagh, 1980) โดยพิจารณาตัวแปรตอบสนองอันดับ (Y) ที่มี k กลุ่ม โดยความเกี่ยวพันระหว่างตัวแปรอธิบาย (x's) กับความน่าจะเป็นสะสมของตัวแปรตอบสนอง Y ดังสมการที่ (1)

$$\log\left[\frac{P(Y \leq j | \mathbf{x})}{P(Y > j | \mathbf{x})}\right] = \alpha_j + \beta' \mathbf{x}, i = 1, \dots, p, j = 1, \dots, k-1 \quad (1)$$

เมื่อ  $\alpha_1 \leq \alpha_2 \leq \dots \leq \alpha_{k-1}$  แทนจุดตัด หรือแทน log odds ของ Y เมื่อ X มีค่าเป็นศูนย์  $\beta'$  แทนเวกเตอร์ของสัมประสิทธิ์ของปัจจัยเสี่ยงหรือตัวแปรอธิบาย ตัวแบบนี้มีข้อสมมติว่า ความเกี่ยวพันของ X และ Y ไม่ขึ้นกับ j หรือการแบ่ง Y ออกเป็นกลุ่ม ๆ และเรียกข้อสมมติของความเท่ากันของ log odds ratio ดังกล่าวว่า Proportional odds assumption ภายใต้ตัวแบบ Proportional odds model ส่วนการประมาณ  $\beta'$  อาศัยวิธีฟิชเชอร์-สกอร์ริง (Fisher scoring method) คำนวณตัวประมาณภาวะน่าจะเป็นสูงสุด (maximum likelihood method) และอาจใช้ตัวสถิติ Wald ในการอนุมานเกี่ยวกับ  $\beta$  ต่อไป

ตัวแบบลอจิสต์ 2 กลุ่ม จะมีรูปแบบเป็นกรณีพิเศษของตัวแบบลอจิสต์ผสม เมื่อศึกษาตัวแปรตอบสนองมีเพียง 2 กลุ่มเท่านั้น ดังสมการที่ (2)

$$\log\left[\frac{P(Y=1 | \mathbf{x})}{1 - P(Y=1 | \mathbf{x})}\right] = \alpha + \beta'\mathbf{x}, \quad (2)$$

เมื่อ  $\alpha$  แทนจุดตัด

ตัวแบบล็อกลิเนียร์ (Birch, 1963; Agresti, 2002) เริ่มจากตารางที่มี 3 ทาง (three-way table) ซึ่งขยายให้เป็นตารางการจรหลายทาง (multi-way contingency table) ได้ เพื่ออธิบายความซับซ้อนเกี่ยวกับอิทธิพลร่วมของตัวแปรในมิติต่าง ๆ และอาศัยฟังก์ชันลอการิทึม สำหรับแปลงค่าคาดหวังของตัวแปรตอบสนองด้วยฟังก์ชันเชื่อมโยงแบบ Log นอกจากนี้การที่มีจำนวนเซลล์ซึ่งจะเพิ่มตามจำนวนตัวแปรและจำนวนระดับตัวแปรนั้น อาจมีผลทำให้บางเซลล์หรือหลายเซลล์มีจำนวนนับเป็นศูนย์ได้ อย่างไรก็ตามปัญหานี้จะลดน้อยลงได้ เมื่อใช้ขนาดตัวอย่างใหญ่มากพอ หรือใช้ตัวแบบล็อกลิเนียร์ แต่ถ้ายังคงมีปัญหาดังกล่าวนี้อยู่อีก ให้ตรวจสอบค่าศูนย์ดังกล่าวว่า มีสาเหตุจากการเลือกตัวอย่าง (sampling zeroes) หรือจากโครงสร้าง (structural zeroes) ซึ่งสามารถวิเคราะห์ด้วยตัวแบบล็อกลิเนียร์ที่เหมาะสมกับปัญหา ตัวอย่างของตัวแบบล็อกลิเนียร์ 5 มิติ เช่น (3)

สมมติมีตัวแปรจำแนกประเภทหรือเชิงกลุ่ม 5 ตัว คือ H P X Y และ Z ตัวแบบของความเป็นอิสระ mutual independence แทนด้วย (H, P, X, Y, Z) ถ้าตัวแบบไม่มีอิทธิพลร่วมของ 3 ตัวแปรซึ่งจะไม่มีอิทธิพลร่วมของ 4 ตัวแปรหรือมากกว่าด้วย จะมีรูปแบบเป็น (HP, HX, HY, HZ, PX, PY, PZ, XY, XZ, YZ) ดังสมการที่ (3)

$$\begin{aligned} \log m_{hijk} = & u + u_{\mathbf{h}}^{\mathbf{H}} + u_{\mathbf{i}}^{\mathbf{P}} + u_{\mathbf{j}}^{\mathbf{X}} + u_{\mathbf{k}}^{\mathbf{Y}} + u_{\mathbf{l}}^{\mathbf{Z}} + u_{\mathbf{hi}}^{\mathbf{HP}} + u_{\mathbf{hj}}^{\mathbf{HX}} + u_{\mathbf{hk}}^{\mathbf{HY}} + u_{\mathbf{jk}}^{\mathbf{XY}} + u_{\mathbf{kl}}^{\mathbf{YZ}} + u_{\mathbf{jl}}^{\mathbf{XZ}} + u_{\mathbf{hl}}^{\mathbf{HZ}} \\ & + u_{\mathbf{ij}}^{\mathbf{PX}} + u_{\mathbf{ik}}^{\mathbf{PY}} + u_{\mathbf{il}}^{\mathbf{PZ}} \end{aligned} \quad (3)$$

เมื่อ  $u$  แทนค่าเฉลี่ยทั้งหมดของ  $\log(m_{hijkl})$  โดยที่  $\{m_{hijkl}\}$  แทนค่าคาดหวังที่สนใจประมาณจากตัวแบบล็อกลิเนียร์ ส่วนเทอมที่เหลือเช่น  $\{u_{\mathbf{h}}^{\mathbf{H}}\}$ ,  $\{u_{\mathbf{i}}^{\mathbf{P}}\}$  และ  $\{u_{\mathbf{hi}}^{\mathbf{HP}}\}$  เปรียบเสมือนอิทธิพลหรือสัมประสิทธิ์ของ H, P และของ HP ตามลำดับ

ตัวแบบ (3) แต่ละคู่ของตัวแปรเป็นอิสระต่อกันแบบมีเงื่อนไข เมื่อกำหนดตัวแปร 2 ตัวอื่น ๆ เช่น  $u_{\mathbf{kl}}^{\mathbf{YZ}} = 0$  ทุก ๆ k และ l แล้ว Y และ Z จะเป็นอิสระต่อกันแบบมีเงื่อนไข ณ แต่ละหมวดหมู่ของระดับของ H และ P ดังนั้นการเป็นอิสระต่อกันแบบมีเงื่อนไขของตัวแปรคู่ใด คือไม่มีเทอมอิทธิพลร่วมของตัวแปรคู่นั้น

ตัวแบบปัวซอง (Poisson models) เป็นตัวแบบที่ตัวแปรสุ่มคือ จำนวนนับที่มีการแจกแจงปัวซอง

ให้เวกเตอร์ตอบสนอง คือ  $\mathbf{Y} = (Y_1, \dots, Y_n)'$  โดยที่  $n$  แทนขนาดของตัวอย่าง และ  $Y_i, Y_j$  มีการแจกแจงเหมือนกันและเป็นอิสระต่อกันสำหรับทุก ๆ  $i \neq j$  โดยฟังก์ชันมวลความน่าจะเป็นของการแจกแจงปัวซองสำหรับตัวแปรสุ่ม  $Y$  (Agresti, 2002) คือ

$$P(y) = \frac{e^{-\mu} \mu^y}{y!}, y = 0, 1, 2, \dots$$

เมื่อ  $E(Y) = \text{Var}(Y) = \mu$ ;  $\mu > 0$  โดยที่ ค่าเฉลี่ยและความแปรปรวนมีค่าเท่ากับ  $\mu$

#### การแจกแจงปัวซองนัยทั่วไป (generalized Poisson distribution)

ในปี 1973 Consul และ Jain ได้เสนอการแจกแจงแบบปัวซองนัยทั่วไปสำหรับการเกิดความแปรปรวนมากกว่าค่าเฉลี่ย (overdispersion) ในข้อมูลจำนวนนับ ซึ่งมีพารามิเตอร์ 2 ตัวคือ  $(\mu, \phi)$  หรือในรูปแบบแกมมาแบบ two gamma function ที่มีฟังก์ชันมวลความน่าจะเป็นของการแจกแจงปัวซองนัยทั่วไปสำหรับตัวแปรสุ่ม  $Y$  คือ

$$P(y; \mu, \phi) = \frac{e^{-(\mu-\phi y)} [\mu(\mu + \phi y)]^{y-1}}{y!}, y = 0, 1, 2, \dots$$

เมื่อ  $\mu$  แทนค่าเฉลี่ย  $\mu > 0$ ,  $\phi$  แทนพารามิเตอร์การกระจาย (dispersion factor)

การแจกแจงปัวซองนัยทั่วไปนี้มีค่าเฉลี่ย  $\mu / (1 - \phi)$  และความแปรปรวน  $\phi E(Y)$  (Yang Hardin and Addy, 2009: อ้างอิงใน Joe and Zhu, 2005)  $\phi = \frac{1}{(1-\phi)^2}$  แทนเทอมการกระจาย

ถ้า  $\phi = 0$  การแจกแจงปัวซองนัยทั่วไปจะลดรูปเป็นการแจกแจงแบบปัวซองที่มีพารามิเตอร์  $\mu$  และถ้า  $\phi < 0$  จะเกิด underdispersion ในตัวแบบ แต่ถ้า  $\phi > 0$  จะเกิด overdispersion ในตัวแบบ

#### ตัวแบบการถดถอยปัวซอง (Poisson regression model)

ตัวแบบที่ใช้ทั่วไปสำหรับข้อมูลจำนวนนับ คือ ตัวแบบการถดถอยปัวซอง และศึกษาความเกี่ยวพันระหว่างตัวแปรตอบสนองที่มีการแจกแจงปัวซองกับตัวแปรอธิบายที่อย่างน้อย 1 ตัวเป็นตัวแปรต่อเนื่อง โดยใช้ฟังก์ชันเชื่อมโยงแบบ log link function ถ้าตัวแปรอธิบายทุกตัวแปรเป็นแบบจำแนกประเภท (categorical variables) ตัวแบบจะสอดคล้องกับตัวแบบล็อกลิเนียร์ข้างต้น ตัวแบบการถดถอยปัวซอง มีรูปแบบดังสมการ (4)

$$\log \mu_i = \mathbf{x}_i \boldsymbol{\beta} \quad (4)$$

เมื่อ  $\mu_i$  แทนค่าคาดหวังแถวที่  $i$ ,  $(1, \mathbf{x}_i)$  แทนแถวที่  $i$  ของเมทริกซ์ของตัวแปรร่วม  $\mathbf{x}$  และ  $\boldsymbol{\beta}$  แทนเวกเตอร์ของพารามิเตอร์ของตัวแบบการถดถอยปัวซองที่ไม่ทราบค่าที่ต้องการประมาณ (Cameron and Trivedi, 1998; Agresti, 2002)

#### ตัวแบบการถดถอยปัวซองนัยทั่วไป (generalize Poisson (GP) regression model)

เป็นตัวแบบที่มี 2 รูปแบบที่ขึ้นอยู่กับลักษณะความเกี่ยวพันระหว่างค่าเฉลี่ยและความแปรปรวน คือ

##### ตัวแบบการถดถอยปัวซองนัยทั่วไปแบบที่ 1 (GP Type I: GP1)

GP1 เป็นตัวแบบที่ค่าความแปรปรวนเป็นฟังก์ชันเชิงเส้นของค่าเฉลี่ย ตัวแบบการถดถอยปัวซองนัยทั่วไปแบบที่ 1 มีฟังก์ชันมวลความน่าจะเป็น ดังนี้

$$P(Y_i = y_i | \mathbf{x}_i, \boldsymbol{\beta}, \phi) = ((1-\phi)\mu_i + \phi y_i)^{y_i-1} \frac{(1-\phi)\mu_i}{y_i!} e^{-(1-\phi)\mu_i - \phi y_i}, y_i = 0, 1, 2, \dots$$

เมื่อ ค่าเฉลี่ยเท่ากับ  $\mu_i = e^{x_i\beta}$  และความแปรปรวนเท่ากับ  $\phi\mu_i$  ซึ่ง  $\phi = \frac{1}{(1-\phi)^2}$  แทนเทอมหรือปัจจัยการกระจาย (dispersion factor) และ  $\phi$  แทน พารามิเตอร์การกระจาย (dispersion parameter) โดยที่  $\phi$  เป็นค่าคงตัว

ค่า  $\phi < 0$  เมื่อความแปรปรวนน้อยกว่าค่าเฉลี่ย ค่า  $\phi > 0$  เมื่อความแปรปรวนมากกว่าค่าเฉลี่ย และเมื่อ  $\phi = 0$  ตัวแบบการถดถอยปัวซองนี้ทั่วไปแบบที่ 1 จะลดรูปเป็นตัวแบบการถดถอยปัวซอง

### ตัวแบบการถดถอยปัวซองนี้ทั่วไปแบบที่ 2 (GP Type II: GP2)

GP2 เป็นตัวแบบที่ค่าความแปรปรวนไม่เป็นฟังก์ชันเชิงเส้นของค่าเฉลี่ย (อยู่ในรูปแบบกำลังสาม) ตัวแบบการถดถอยปัวซองนี้ทั่วไปแบบที่ 2 มีฟังก์ชันมวลความน่าจะเป็นของการแจกแจงในรูปแบบ คือ

$$P(Y_i = y_i | x_i, \beta, \phi) = \left( \frac{\mu_i}{1 + \phi\mu_i} \right)^{y_i} \frac{(1 + \phi\mu_i)^{y_i-1}}{y_i!} e^{-\frac{\mu_i(1+\phi\mu_i)}{1+\phi\mu_i}}, \quad y_i = 0, 1, 2, \dots$$

เมื่อ ค่าเฉลี่ยเท่ากับ  $\mu_i = e^{x_i\beta}$  และความแปรปรวนคือ  $\mu_i(1 + \phi\mu_i)^2$  ส่วน  $\phi$  คือพารามิเตอร์การกระจาย และ  $\phi$  เป็นค่าคงตัว โดย  $\phi > \min \left( -\frac{1}{\max(\mu_i)}, -\frac{1}{\max(y_i)} \right)$

เมื่อ  $\phi < 0$  จะทำให้ความแปรปรวนน้อยกว่าค่าเฉลี่ย และเรียกกรณีนี้ว่า กรณี Underdispersion

เมื่อ  $\phi > 0$  จะทำให้ความแปรปรวนมากกว่าค่าเฉลี่ย และเรียกกรณีนี้ว่า กรณี Overdispersion

เมื่อ  $\phi = 0$  ตัวแบบการถดถอยปัวซองนี้ทั่วไปแบบที่ 2 จะลดรูปเป็นตัวแบบการถดถอยปัวซอง

ตัวแบบ GP1 และ GP2 ขยายจากตัวแบบการถดถอยปัวซอง โดยตัวแบบ GP1 มีความแปรปรวนที่เป็นฟังก์ชันเชิงเส้นของค่าเฉลี่ย ส่วนตัวแบบ GP2 มีความแปรปรวนที่เป็นฟังก์ชันไม่เชิงเส้นของค่าเฉลี่ยในรูปแบบกำลังสาม (Yang, Hardin & Addy, 2009) และประมวลผลข้อมูลด้วยโปรแกรม SAS Version 9.3 และ SAS Enterprise 5.3

## ผลการวิจัย (results)

ผลการวิจัยภายใต้วิธีขั้นตอน (stepwise method) เป็น 4 กรณี

### กรณีที่ 1: ผลการวิจัยโดยใช้ตัวแบบลอจิสติก

การทดสอบตัวแบบลอจิสติกแบบ proportional odds พบว่า สอดคล้องกับข้อสมมติของตัวแบบที่  $\alpha = 0.05$  (P-value = 0.3112) และปัจจัยเสี่ยงที่ถูกคัดเลือกด้วยวิธีอัตราส่วนภาวะน่าจะเป็น (stepwise likelihood ratio: LR) ได้แก่ Ptl (การคลอดก่อนกำหนด), Ht (ประวัติโรคความดันโลหิตสูงของมารดา และ Ui (การมีพยาธิสรีดวะ) อย่างมีนัยสำคัญทางสถิติ ที่ 0.05 (P-value = 0.0089, 0.0060, 0.0179) ตามลำดับ การทดสอบและค่าสัมประสิทธิ์ของตัวแบบดังตารางที่ 1

ค่าประมาณตัวแบบลอจิสติก ดัง (5)-(6)

$$\text{Logit } [P(Y \leq 1 | \mathbf{x}_i)] = -1.0297 - 0.5225 \text{ Ptl } (0) + 0.8019 \text{ Ptl } (1) - 0.7923 \text{ Ht} - 0.4989 \text{ Ui} \quad (5)$$

$$\text{Logit } [P(Y \leq 2 | \mathbf{x}_i)] = 0.5315 - 0.5225 \text{ Ptl } (0) + 0.8019 \text{ Ptl } (1) - 0.7923 \text{ Ht} - 0.4989 \text{ Ui} \quad (6)$$



ตารางที่ 1 ค่าประมาณสัมประสิทธิ์ของตัวแบบลอจิสติกส์ เมื่อตัวแปรตอบสนองมี 3 ระดับ

| พารามิเตอร์ | กลุ่ม | DF | $\hat{\beta}$ | SE ( $\hat{\beta}$ ) | สถิติ Wald | P-Value |
|-------------|-------|----|---------------|----------------------|------------|---------|
| Intercept   | 1     | 1  | -1.0297       | 0.4386               | 5.5116     | 0.0189  |
| Intercept   | 2     | 1  | 0.5315        | 0.4248               | 1.5658     | 0.2108  |
| Ptl         | 0     | 1  | -0.5225       | 0.3392               | 2.3723     | 0.1235  |
| Ptl         | 1     | 1  | 0.8019        | 0.3874               | 4.2848     | 0.0385  |
| Ht          | 0     | 1  | -0.7923       | 0.2884               | 7.5491     | 0.006   |
| Ui          | 0     | 1  | -0.4989       | 0.2106               | 5.6094     | 0.0179  |

ตัวอย่างการประมาณค่าความน่าจะเป็นของตัวแปรตอบสนอง ดังนี้

ค่าความน่าจะเป็นที่หญิงจะคลอดบุตรอยู่ในระดับน้ำหนักร้อยที่เสี่ยงต่อการเสียชีวิต คือ  $P(Y=1) = P(Y \leq 1)$  ค่าความน่าจะเป็นที่หญิงจะคลอดบุตรอยู่ในระดับน้ำหนักร้อยที่ต้องเฝ้าระวังคือ  $P(Y=2) = P(Y \leq 2) - P(Y \leq 1)$  ค่าความน่าจะเป็นที่หญิงจะคลอดบุตรอยู่ในระดับน้ำหนักปกติ คือ  $P(Y=3) = 1 - P(Y \leq 2)$  ตัวอย่างการวิเคราะห์ข้อมูลและการตีความหมายของผลลัพธ์ เช่น ถ้าต้องการพยากรณ์หญิงตั้งครรภ์คนหนึ่งซึ่งเป็นโรคความดันโลหิตสูงว่าจะมีค่าความน่าจะเป็นที่หญิงตั้งครรภ์จะคลอดบุตรโดยที่ระดับน้ำหนักของทารกแรกเกิด ในระดับน้ำหนักร้อยเสี่ยงต่อการเสียชีวิต คือ

$$P(Y=1) = P(Y \leq 1) = \frac{e^{-1.0297 - 0.5225(0) + 0.8019(0) - 0.7923(1) - 0.4989(0)}}{1 + e^{-1.0297 - 0.5225(0) + 0.8019(0) - 0.7923(1) - 0.4989(0)}} = 0.1392$$

ดังนั้นถ้าหญิงตั้งครรภ์เป็นโรคความดันโลหิตสูง ความน่าจะเป็นที่หญิงตั้งครรภ์จะคลอดบุตรที่มีน้ำหนักแรกเกิดอยู่ในระดับเสี่ยงต่อการเสียชีวิต เท่ากับ 0.1392 หรือในระดับสูงถึง 13% และค่าความน่าจะเป็นที่หญิงจะคลอดบุตรที่ระดับน้ำหนักของทารกแรกเกิด คือ ระดับน้ำหนักร้อยที่ต้องเฝ้าระวัง คือ

$$P(Y=2) = P(Y \leq 2) - P(Y \leq 1) = 0.4352 - 0.1392 = 0.2960$$

ดังนั้น ถ้าหญิงตั้งครรภ์เป็นโรคความดันโลหิตสูง ความน่าจะเป็นที่หญิงตั้งครรภ์จะคลอดบุตร ที่มีน้ำหนักแรกเกิดอยู่ในระดับน้ำหนักร้อยที่ต้องเฝ้าระวัง เท่ากับ 0.2960 หรือในระดับสูงถึงเกือบ 30% ส่วนค่าความน่าจะเป็นที่หญิงตั้งครรภ์จะคลอดบุตรโดยที่ระดับน้ำหนักของทารกแรกเกิดปกติ คือ  $P(Y=3) = 1 - P(Y \leq 2) = 1 - 0.4352 = 0.5648$  ดังนั้น ถ้าหญิงตั้งครรภ์เป็นโรคความดันโลหิตสูง ความน่าจะเป็นที่หญิงตั้งครรภ์จะคลอดบุตรที่มีน้ำหนักแรกเกิดอยู่ในระดับน้ำหนักปกติ เท่ากับ 0.5648 หรือในระดับเพียง 56.48% นอกจากนี้ความน่าจะเป็นที่หญิงตั้งครรภ์จะคลอดบุตรที่ระดับน้ำหนักของทารกแรกเกิดระดับอื่น ๆ จะทำได้ทำนองเดียวกัน จากตารางที่ 1 การประมาณและผลลัพธ์ odds ratio ดังตารางที่ 2

เมื่อพิจารณากลุ่มทารกแรกเกิดที่มีน้ำหนักอยู่ในระดับเสี่ยงต่อการเสียชีวิต หรือระดับที่ต้องเฝ้าระวังเปรียบเทียบกับจำนวนทารกที่มีระดับปกติ ที่  $\alpha = 0.05$  (ตารางที่ 2) พบว่า

ตารางที่ 2 ค่าประมาณแบบจุดและแบบช่วงสำหรับ odds ratio กรณีตัวแบบลอจิสติก

| ปัจจัยเสี่ยง | ค่าสัมประสิทธิ์ | 95% ช่วงความเชื่อมั่น (95% Wald CI) |
|--------------|-----------------|-------------------------------------|
| Ptl 0 vs 2   | 0.784           | 0.141 4.375                         |
| Ptl 1 vs 2   | 2.949           | 1.472 18.429                        |
| Ht 0 vs 1    | 0.205           | 0.066 0.635                         |
| Ui 0 vs 1    | 0.369           | 0.161 0.842                         |

- หญิงที่ไม่ได้คลอดก่อนกำหนด มีโอกาสคลอดบุตรที่มีระดับน้ำหนักเสี่ยงต่อการเสียชีวิต น้อยเป็น 0.784 เท่า ของหญิงที่คลอดก่อนกำหนดมากกว่า 1 เดือน กล่าวอีกนัยหนึ่ง หญิงที่คลอดก่อนกำหนดมากกว่า 1 เดือน มีโอกาสคลอดบุตรที่มีระดับน้ำหนักเสี่ยงต่อการเสียชีวิต มากเป็น 1.275 เท่าของหญิงที่ไม่ได้คลอดก่อนกำหนด โดยในช่วง 95% ของ Wald กลุ่ม 1 (0.141, 4.375) ความเสี่ยงจึงไม่ต่างกันที่ระดับนัยสำคัญ 0.05

- หญิงที่คลอดก่อนกำหนดเป็นระยะเวลาภายใน 1 เดือน มีโอกาสคลอดบุตรที่มีระดับน้ำหนักเสี่ยงต่อการเสียชีวิต เป็น 2.949 เท่า ของหญิงที่คลอดก่อนกำหนดมากกว่า 1 เดือนโดยให้ช่วงความเชื่อมั่น 95% ไม่คลุม 1 (1.472, 18.429) ความเสี่ยงดังกล่าวจึงแตกต่างกันอย่างมีนัยสำคัญทางสถิติ 0.05 การตีความหมายของผลลัพธ์อื่นๆ ทำได้ทำนองเดียวกัน

### กรณีที่ 2: ผลการวิจัยโดยใช้ตัวแบบลอจิสติก 2 กลุ่ม

ผลลัพธ์การวิเคราะห์พบว่า ตัวแบบเหมาะสมและมีภาวะสารูปดี โดยปฏิเสธความเป็นอิสระที่ระดับนัยสำคัญทางสถิติ 0.05 ( $P=0.0113$ ) และปัจจัยเสี่ยงที่มีนัยสำคัญทางสถิติที่  $\alpha =0.05$  ได้แก่ การคลอดก่อนกำหนด (Ptl,  $P=0.0102$ ) ประวัติโรคความดันโลหิตสูงของมารดา (Ht,  $P=0.0261$ ) และมดลูกถูกกระทบกระเทือน (Ui,  $P=0.0324$ ) ดังในตารางที่ 3

ตารางที่ 3 ค่าประมาณสัมประสิทธิ์ของตัวแบบลอจิสติก 2 กลุ่ม

| พารามิเตอร์ | น้ำหนักเด็ก | DF | $\hat{\beta}$ | SE ( $\hat{\beta}$ ) | ค่าสถิติ Wald | P-value |
|-------------|-------------|----|---------------|----------------------|---------------|---------|
| Intercept   |             | 1  | 0.4799        | 0.4591               | 1.0927        | 0.2959  |
| Ptl         | 0           | 1  | -0.5909       | 0.3603               | 2.689         | 0.101   |
| Ptl         | 1           | 1  | 1.0954        | 0.4264               | 6.5996        | 0.0102  |
| Ht          | 0           | 1  | -0.7079       | 0.3182               | 4.9498        | 0.0261  |
| Ui          | 0           | 1  | -0.4832       | 0.2258               | 4.5797        | 0.0324  |

ค่าประมาณและตัวแบบลอจิสติก 2 กลุ่ม ดัง (7)

$$\text{Logit } [P(Y = 1 \mid \mathbf{x}_i)] = 0.4799 - 0.5909 \text{ Ptl } (0) + 1.0954 \text{ Ptl } (1) - 0.7079 \text{ Ht} - 0.4832 \text{ Ui} \quad (7)$$

### ตัวอย่างผลลัพธ์การตีความหมายของผลลัพธ์ด้วยตัวแบบลอจิสติก 2 กลุ่ม

ถ้าต้องการพยากรณ์หญิงตั้งครรภ์คนหนึ่งซึ่งเป็นโรคความดันโลหิตสูง ว่าจะมีค่าความน่าจะเป็นที่หญิงตั้งครรภ์จะคลอดบุตรทารกที่มีระดับน้ำหนักเสี่ยงต่อการเสียชีวิต ทำได้ดังนี้

$$P(Y=1) = \frac{e^{0.4799 - 0.5909 \text{ Ptl}(0) + 1.0954 \text{ Ptl}(0) - 0.7079 (1) - 0.4832 (0)}}{1 + e^{0.4799 - 0.5909 \text{ Ptl}(0) + 1.0954 \text{ Ptl}(0) - 0.7079 (1) - 0.4832 (0)}} = 0.4432$$

ดังนั้น ถ้าหญิงตั้งครรภ์เป็นโรคความดันโลหิตสูง ความน่าจะเป็นที่หญิงตั้งครรภ์จะคลอดบุตร ที่มีน้ำหนักแรกเกิดอยู่ในระดับเสี่ยงต่อการเสียชีวิต เท่ากับ 0.4432 หรือสูงประมาณ 44.32% ความน่าจะเป็นที่หญิงจะคลอดทารกแรกเกิดที่ระดับน้ำหนักปกติคือ  $P(Y=2) = 1 - P(Y=1) = 1 - 0.4432 = 0.5568$

ถ้าหญิงตั้งครรภ์เป็นโรคความดันโลหิตสูง ความน่าจะเป็นที่หญิงตั้งครรภ์จะคลอดบุตร ที่มีน้ำหนักแรกเกิดอยู่ในระดับน้ำหนักปกติ เท่ากับ 0.5568 (หรือ 56.68%) ส่วนการพยากรณ์หญิงตั้งครรภ์ด้วยตัวแปรอธิบายอื่นๆ ทำได้ในทำนองเดียวกัน

การตีความหมายของผลลัพธ์และ odds ratio กรณีตัวแบบลอจิสต์ 2 กลุ่ม (ตารางที่ 4) ตัวอย่างเช่น

1) หญิงที่ไม่ได้คลอดก่อนกำหนด มีโอกาสคลอดบุตรที่มีระดับน้ำหนักเสี่ยงต่อการเสียชีวิต น้อยเป็น 0.917 เท่าของหญิงที่คลอดก่อนกำหนดมากกว่า 1 เดือน กล่าวอีกนัยหนึ่ง หญิงที่คลอดก่อนกำหนดภายใน 1 เดือน มีโอกาสคลอดบุตรที่มีน้ำหนักในระดับเสี่ยงต่อการเสียชีวิต มากเป็น 1.091 เท่าของหญิงที่ไม่ได้คลอดก่อนกำหนด แต่จากช่วงความเชื่อมั่น 95% ของ Odds ratio กลุ่ม 1 (0.149, 5.65) จึงไม่พบความแตกต่างอย่างมีนัยสำคัญที่ 0.05

2) หญิงตั้งครรภ์ซึ่งไม่เป็นโรคความดันโลหิตสูง มีโอกาสคลอดบุตรที่มีระดับน้ำหนักเสี่ยงต่อการเสียชีวิต น้อยเป็น 0.243 เท่าของหญิงซึ่งเป็นโรคความดันโลหิตสูง หรือหญิงที่เป็นโรคความดันโลหิตสูง มีโอกาสคลอดบุตรที่มีระดับน้ำหนักเสี่ยงต่อการเสียชีวิต มากเป็น 4.115 เท่าของหญิงซึ่งไม่เป็นโรคความดันโลหิตสูง ( $1/0.2434$ ) อย่างมีนัยสำคัญทางสถิติที่  $\alpha = 0.05$

ตารางที่ 4 ค่าประมาณสำหรับ odds ratio กรณีตัวแบบลอจิสต์ 2 กลุ่ม

| Effect     | Point Estimate | 95 Wald % CI |
|------------|----------------|--------------|
| Ptl 0 vs 2 | 0.917          | 0.149 5.65   |
| Ptl 1 vs 2 | 4.953          | 0.687 35.725 |
| Ht 0 vs 1  | 0.243          | 0.07 0.845   |
| Ui 0 vs 1  | 0.38           | 0.157 0.922  |

### กรณีที่ 3: ผลการวิจัยโดยใช้ตัวแบบล็อกลิเนียร์

จากกรณีที่ 2 ซึ่งปัจจัยที่พบคือ ประวัติโรคความดันโลหิตสูงของมารดา การคลอดก่อนกำหนด และ น้ำหนักของทารกแรกเกิด นำมาศึกษาด้วยตัวแบบล็อกลิเนียร์ที่มีเทอมของอิทธิพลร่วมเพิ่มเติมได้ผลดังนี้

$$\log(m_{ijk}) = u + u_i^{ht} + u_j^{ui} + u_k^{bwt} + u_{ij}^{ht*ui} + u_{ik}^{ht*bwt} + u_{jk}^{ui*bwt} + u_{ijk}^{ht*ui*bwt} \quad (8)$$

$$\log(m_{ijk}) = u + u_i^{ht} + u_j^{ui} + u_k^{bwt} + u_{ij}^{ht*ui} + u_{ik}^{ht*bwt} + u_{jk}^{ui*bwt}$$

$$\log(m_{ijk}) = u + u_i^{ht} + u_j^{ui} + u_k^{bwt} + u_{ij}^{ht*ui} + u_{ik}^{ht*bwt}$$

$$\log(m_{ijk}) = u + u_i^{ht} + u_j^{ui} + u_k^{bwt} + u_{ij}^{ht*ui}$$

$$\log(m_{ijk}) = u + u_i^{ht} + u_j^{ui} + u_k^{bwt}$$

ตัวแบบเต็ม (saturated model) คือ (8) ถัดลงมาคือ ตัวแบบไม่เต็ม (unsaturated models) ต่าง ๆ ซึ่งมีพารามิเตอร์น้อยกว่า และสนใจตรวจสอบว่าเป็นตัวแบบเหมาะสม เพื่อช่วยในการวิเคราะห์ข้อมูลต่อไป

**ตัวอย่างผลลัพธ์ของค่าประมาณพารามิเตอร์ของตัวแบบเต็มลือกลิเนียร์ ดังนี้**

ตัวแบบลือกลิเนียร์ สำหรับเซลล์ (0,0,1)

$$\log(m_{001}) = u + u_0^{ht} + u_0^{ui} + u_1^{bwt} + u_{00}^{ht*ui} + u_{01}^{ht*bwt} + u_{01}^{ui*bwt} + u_{001}^{ht*ui*bwt}$$

$$\log(m_{001}) = -21.6931 + 23.4848 + 23.0794 - 0.0001 - 22.6739 + 0.8474 + 0.2233 + 1.4417 = 4.7095$$

ตัวแบบลือกลิเนียร์ สำหรับเซลล์ (0,0,2)

$$\log(m_{002}) = u + u_0^{ht} + u_0^{ui} + u_2^{bwt} + u_{00}^{ht*ui} + u_{02}^{ht*bwt} + u_{02}^{ui*bwt} + u_{002}^{ht*ui*bwt}$$

$$\log(m_{002}) = -21.6931 + 23.4848 + 23.0794 - 0.0001 - 22.6739 + 0.2878 - 0.2876 + 1.1700 = 3.3673$$

ตัวแบบลือกลิเนียร์ สำหรับเซลล์ (1,1,3)

$$\log(m_{113}) = u + u_1^{ht} + u_1^{ui} + u_3^{bwt} + u_{11}^{ht*ui} + u_{13}^{ht*bwt} + u_{13}^{ui*bwt} + u_{113}^{ht*ui*bwt}$$

$$\log(m_{113}) = -21.6931 + 0 + 0 + 0 + 0 + 0 + 0 + 0 = -21.6931$$

**ตัวอย่างผลลัพธ์ของตัวประมาณพารามิเตอร์ของตัวแบบไม่เต็ม เช่น**

$$\log(m_{ijk}) = u + u_i^{ht} + u_j^{ui} + u_k^{bwt} + u_{ij}^{ht*ui} \text{ ซึ่งให้ผลลัพธ์ เช่น}$$

ตัวแบบลือกลิเนียร์ สำหรับเซลล์ (0,0,1)

$$\log(m_{001}) = u + u_0^{ht} + u_0^{ui} + u_1^{bwt} + u_{00}^{ht*ui}$$

$$\log(m_{001}) = -23.6290 + 24.6639 + 23.8166 + 1.9231 - 22.1448 = 4.6298$$

การเลือกตัวแบบที่เหมาะสมระหว่างตัวแบบเต็มและไม่เต็มต่าง ๆ อาศัยค่าผลต่างของ deviances ระหว่างตัวแบบที่เปรียบเทียบ เพื่อหาตัวแบบที่เหมาะสมมากกว่าพบว่า ตัวแบบใน (9) เหมาะสมที่สุดอย่างมีนัยสำคัญทางสถิติ ( $P = 0.9824$ ) โดยวิธีการเลือกแบบขั้นตอน นั่นคือ

$$\log(m_{ijk}) = u + u_i^{ht} + u_j^{ui} + u_k^{bwt} + u_{ij}^{ht*ui} + u_{ik}^{ht*bwt} + u_{jk}^{ui*bwt} \quad (9)$$

จากตัวแบบ (9) สามารถสรุปค่าสังเกต ( $o_{ijk}$ ) และค่าประมาณของค่าคาดหวัง ( $m_{ijk}$ ) ของแต่ละเซลล์ ไว้ในตารางที่ 5

ตารางที่ 5 ค่าสังเกต ( $O_{ijk}$ ) และค่าประมาณของค่าคาดหวัง ( $m_{ijk}$ ) ในตารางการจร (2x2x3)

| ประวัติ<br>ความดัน<br>โลหิตสูง | การมี<br>ปัญหา<br>สูตินารี<br>เวช | น้ำหนักทารกแรกเกิด (กรัม)                            |                                                      |                                                      |
|--------------------------------|-----------------------------------|------------------------------------------------------|------------------------------------------------------|------------------------------------------------------|
|                                |                                   | ระดับปกติ                                            | ระดับต้องเฝ้าระวัง                                   | ระดับเสี่ยงต่อการเสียชีวิต                           |
| 1 (Yes)                        | 1 (Yes)                           | $O_{111} = 0,$<br>$m_{111} = 7.3738 \times 10^{-11}$ | $O_{112} = 0,$<br>$m_{112} = 9.6768 \times 10^{-11}$ | $O_{113} = 0,$<br>$m_{113} = 3.1182 \times 10^{-10}$ |
| 1 (Yes)                        | 0 (No)                            | $O_{101} = 5,$<br>$m_{101} = 4.9999$                 | $O_{102} = 3,$<br>$m_{102} = 2.9999$                 | $O_{103} = 4,$<br>$m_{103} = 4.0000$                 |
| 0 (No)                         | 1 (Yes)                           | $O_{011} = 14,$<br>$m_{011} = 13.9992$               | $O_{012} = 8,$<br>$m_{012} = 7.9997$                 | $O_{013} = 6,$<br>$m_{013} = 5.9996$                 |
| 0 (No)                         | 0 (No)                            | $O_{001} = 111,$<br>$m_{001} = 110.9966$             | $O_{002} = 29,$<br>$m_{002} = 29.0001$               | $O_{003} = 9,$<br>$m_{003} = 8.9998$                 |

ค่าประมาณของค่าคาดหวัง ( $m_{ijk}$ ) นำไปสู่การคำนวณค่า odds และ odds ratio ต่าง ๆ ทั้งเซลล์ที่ค่าสังเกตเป็น 0 และไม่เป็น 0 การตีความหมายของความเสี่ยงจากค่า odds และ odds ratio จากตัวแบบล็อกลิเนียร์ เช่น  $\hat{\theta} = \frac{m_{103}}{m_{102}} = \frac{4}{2.9999} = 1.3334$  หมายความว่า มารดาที่มีประวัติความดันโลหิตสูงและไม่มีปัญหาสูตินารีเวช มีโอกาสที่จะคลอดทารกแรกเกิดที่มีน้ำหนักอยู่ในระดับที่เสี่ยงต่อการเสียชีวิตเป็น 1.3334 เท่าของทารกแรกเกิดที่มีน้ำหนักอยู่ในระดับที่ต้องเฝ้าระวัง

#### กรณี 4: ผลการวิจัยโดยใช้ตัวแบบการถดถอยปัวซองเปรียบเทียบกับตัวแบบการถดถอยปัวซองน้อยทั่วไป

การประมาณค่าสัมประสิทธิ์ในสำหรับตัวแบบการถดถอยปัวซอง (ตารางที่ 6) พบว่า อายุของมารดานัยสำคัญที่ 0.05 อยู่ในตัวแบบการถดถอยปัวซอง นั่นคือ  $\hat{\mu} = 2.5563 - 0.06908x$  เมื่อ  $\hat{\mu}$  แทนค่า Natural logarithm ของเฉลี่ยของจำนวนทารกที่มีน้ำหนักแรกเกิดน้อยที่เสี่ยงต่อการเสียชีวิต และ  $x$  แทนอายุมารดา และในตัวแบบการถดถอยปัวซองน้อยทั่วไปแบบที่ 2 (ตารางที่ 7) พบว่า  $\hat{\phi} = 0.09367$  มีค่าสถิติ  $t = 1.24$  และค่า P-value = 0.2254 ซึ่งมากกว่า 0.05 จึงยอมรับว่าข้อมูลชุดนี้ไม่เกิด Over dispersion และปัจจัยที่มีผลต่อการคลอดบุตรที่มีระดับน้ำหนักเสี่ยงต่อการเสียชีวิต อธิบายด้วยปัจจัยอายุของมารดา (ปี) อย่างมีนัยสำคัญทางสถิติ 0.05 ตัวแบบจึงเหมาะสม

ตารางที่ 6 ค่าประมาณสัมประสิทธิ์ของตัวแบบการถดถอยปัวซอง

| พารามิเตอร์     | DF | $\hat{\beta}$ | SE ( $\hat{\beta}$ ) | ค่าสถิติ t | P-value |
|-----------------|----|---------------|----------------------|------------|---------|
| $\hat{\beta}_0$ | 24 | 2.5563        | 0.4584               | 5.58       | <0.0001 |
| $\hat{\beta}_1$ | 24 | -0.06908      | 0.01971              | -3.51      | 0.0018  |

ตารางที่ 7 ค่าประมาณสัมประสิทธิ์ของตัวแบบการถดถอยปัวซองน้อยทั่วไปแบบที่ 2

| พารามิเตอร์     | DF | $\hat{\beta}$ | SE ( $\hat{\beta}$ ) | ค่าสถิติ t | P-value |
|-----------------|----|---------------|----------------------|------------|---------|
| $\hat{\beta}_0$ | 24 | 2.9468        | 0.743                | 3.97       | 0.0006  |
| $\hat{\beta}$   | 24 | -0.0852       | 0.0348               | -2.8       | 0.01    |
| $\hat{\phi}$    | 24 | 0.09367       | 0.07527              | 1.24       | 0.2254  |

ส่วนผลลัพธ์การพยากรณ์ความน่าจะเป็นของกลุ่มเสี่ยง อาจยกตัวอย่างด้วยตัวแบบลอจิสติกส์ซึ่งพบว่า ถ้าแม่ของทารกมีปัจจัยเสี่ยงที่สอดคล้องกับตัวแบบแล้ว ความน่าจะเป็นของการคลอดบุตรที่มีระดับน้ำหนักเสี่ยงต่อการเสียชีวิตและที่ต้องเฝ้าระวังนั้นมีค่อนข้างสูง

## วิจารณ์ผลการวิจัย (discussion)

กรณีตัวแบบลอจิสติกส์ภายใต้ตัวแบบ proportional odds model พบว่าปัจจัยที่ถูกคัดเลือกด้วยวิธี stepwise ประกอบด้วย การคลอดก่อนกำหนด ประวัติโรคความดันโลหิตสูงของมารดา และการมีปัญหาลูติหารีเวช อย่างมีนัยสำคัญทางสถิติที่ 0.05 ( $P < 0.05$ ) กรณีตัวแบบลอจิสติกส์ 2 กลุ่ม พบว่าปัจจัยที่ได้ตรงกับของกรณีตัวแบบลอจิสติกส์ 3 ระดับ กรณีตัวแบบล็อกลิเนียร์ พบว่า ตัวแปรที่เหมาะสมประกอบด้วย ประวัติการเป็นโรคความดันโลหิตสูงของมารดา การคลอดก่อนกำหนด และระดับน้ำหนักของทารกแรกเกิด และกรณีตัวแบบการถดถอยปัวซงเทียบกับตัวแบบการถดถอยปัวซงน้อยทั่วไปแบบที่ 2 พบว่าค่าประมาณพารามิเตอร์ต่าง ๆ ของตัวแบบการถดถอยปัวซงและตัวแบบการถดถอยปัวซงน้อยทั่วไปแบบที่ 2 ไม่มีปัญหาการเกิด overdispersion จึงเลือกตัวแบบการถดถอยปัวซง และพบตัวแปรอีกตัวหนึ่งคือ อายุของมารดาที่ตั้งครรภ์ มีนัยสำคัญทางสถิติที่ 0.05 ดังนั้นผลการศึกษาด้วยตัวแบบเชิงเส้นน้อยทั่วไป จึงช่วยให้อธิบายปัจจัยที่เสี่ยงต่อการเสียชีวิตของทารกที่พิจารณาจากน้ำหนักแรกเกิดของทารกแรกเกิดหลายปัจจัย อย่างมีนัยสำคัญทางสถิติที่ 0.05 เกือบทุกปัจจัย ทุกตัวแบบที่ใช้มีภาวะสารูปดีกับข้อมูลที่ศึกษาด้วย อย่างไรก็ตามในการวิจัยนี้ได้ศึกษาปัจจัยเสี่ยง จากวรรณกรรมที่เกี่ยวข้อง ในการวิจัยครั้งต่อไปอาจศึกษาปัจจัยเสี่ยงอื่น ๆ ซึ่งอาจเกิดขึ้นในเฉพาะท้องถิ่นหรือที่ผู้เชี่ยวชาญแนะนำเพิ่มเติม มาศึกษาวิเคราะห์เพิ่มเติมได้

## สรุปผลการวิจัย (conclusion)

ผลสรุปจากการวิเคราะห์ปัจจัยเสี่ยงต่อการเสียชีวิตของทารกที่พิจารณาจากน้ำหนักแรกเกิดของทารกด้วยตัวแบบเชิงเส้นน้อยทั่วไป พบว่าตัวแบบทั้ง 4 มีภาวะสารูปดีกับข้อมูลที่ระดับนัยสำคัญ 0.05 และปัจจัยเสี่ยงต่อการเสียชีวิตที่มีผลกระทบต่อน้ำหนักของทารกแรกเกิดอย่างมีนัยสำคัญทางสถิติที่ 0.05 ได้แก่ การคลอดก่อนกำหนด ประวัติโรคความดันโลหิตสูงของมารดา การมีปัญหาลูติหารีเวช และอายุของมารดาที่ตั้งครรภ์ เมื่อนำตัวแบบและปัจจัยเสี่ยงดังกล่าวไปพยากรณ์ เช่นจากตัวแบบลอจิสติกส์ 2 กลุ่ม พบว่าถ้าแม่ของทารกมีปัจจัยเสี่ยงดังกล่าวแล้ว ความน่าจะเป็นของการคลอดบุตรที่มีระดับน้ำหนักเสี่ยงต่อการเสียชีวิตและที่ต้องเฝ้าระวังนั้นมีค่อนข้างสูง

## เอกสารอ้างอิง (references)

- เมษิยา แยมเจริญกิจ. (2554). การทดสอบพารามิเตอร์การกระจายของตัวแบบการถดถอยปัวซงน้อยทั่วไป. วิทยานิพนธ์วิทยาศาสตร์ มหาบัณฑิต มหาวิทยาลัยศิลปากร. 87 หน้า.
- วีรานันท์ พงศาภักดี. (2555). การวิเคราะห์ข้อมูลจำแนกประเภท: ทฤษฎีและการประยุกต์ ด้วย GLIM, SPSS, SAS และ MTB พิมพ์ครั้งที่ 3. นครปฐม: โครงการตำรา คณะวิทยาศาสตร์ มหาวิทยาลัยศิลปากร โรงพิมพ์มหาวิทยาลัยศิลปากร. หน้า 165-180.

- บุญสนอง ภิบุญญ. (2550). การศึกษาปัจจัยที่มีความสัมพันธ์การเกิดทารกแรกเกิดน้ำหนักน้อยกว่า 2,500 กรัม ในเขต 9 แหล่งข้อมูล:  
[http://hpc9.anamai.moph.go.th/research/index.php?option=com\\_content&task=category&sectionid=11&id=28&Itemid=53](http://hpc9.anamai.moph.go.th/research/index.php?option=com_content&task=category&sectionid=11&id=28&Itemid=53) ค้นเมื่อวันที่ 20 มกราคม 2556
- <http://guru.sanook.com> ค้นเมื่อวันที่ 20 มกราคม 2556
- <http://www.anamai.moph.go.th/main.php?filename=index> สืบค้น มกราคม 2556
- Agresti, A. (2002). *Categorical Data Analysis*. 2<sup>nd</sup> ed. New York: John Wiley & Sons. pp. 115–155.
- Birch, M.W. (1963). Maximum likelihood in three-way contingency tables. *J. Roy. Statist. Soc. B* 25: 220-233.
- Cameron, A.C. and Trivedi, P.K. (1998). *Regression Analysis of Count Data*. Cambridge, C, University Press. 411 pages
- Hosmer, D.W. and Lemeshow, S. (1989). *Applied Logistic Regression*. New York: Wiley & Sons. pp. 247–252.
- Joe, H. and Zhu, R. (2005). Generalized Poisson distribution: the property of mixture of Poisson and comparison with negative binomial distribution. *Biometrical Journal* 47: 219–229 Issue 2.
- McCullagh, P. (1980). Regression models for ordinal data. *J. Royal, Statist. Soc. Ser. B* 42: 109-142.
- Schwarz, G. (1978). Estimating the dimensions of a model. *Annals of Statistics* 6: 461-464.
- Walker, S.H. and Duncan, D. B. (1967). Estimation of the probability of event as a function of several independent variables. *Biometrika* 54: 167-179.
- Yang, Z., Hardin, J. W., and Addy, C. L. (2009). A score test for overdispersion in Poisson regression based on the generalized Poisson-2 model. *Journal of Statistical Planning and Inference* 139: 1514-1521.

