ว.วิทย. มข. 41(3) 744-755 (2556) KKU Sci. J. 41(3) 744-755 (2013)

การวิเคราะห์ปัจจัยที่มีผลกระทบต่อการปนเปื้อนของสารหนูในเล็บคน ด้วยตัวแบบล็อกลิเนียร์

Analysis of Factors Affecting the Arsenic Contamination in Toenails Using Loglinear Models

เอกพันธ์ หวานใจ[1] *วีรานันท์ พงศาภักดี*[1*] *และ* *ไพโรจน์ ขาวสิทธิวงศ์*[1]

บทคัดย่อ

การปนเปื้อนสารหนู (arsenic) ในคนอาจมีอันตรายรุนแรงหลายด้านจนอาจเกิดเป็นโรคหรือถึงขั้นเสียชีวิตได้ การปนเปื้อนมีหลายทางเช่น ผิวหนัง เล็บมือ เล็บเท้า น้ำดื่ม ยา หรือ อาหาร งานวิจัยทางสาธารณสุขและทางระบาดวิทยาจึงมีการตรวจสอบในด้านนี้ เพื่อเป็นการเฝ้าระวังและการพยากรณ์ปัจจัยที่มีผลกระทบต่อการปนเปื้อนสารหนูในคน งานวิจัยนี้สนใจปัจจัยที่มีผลกระทบต่อการปนเปื้อนหรือการสะสมของสารหนูในเล็บคน เนื่องจากการปนเปื้อนสารหนูจากแหล่งต่าง ๆ เกิดการสะสมในเล็บได้ โดยศึกษาปัจจัยจำแนกประเภทและแบบเชิงกลุ่ม (categorical and grouped variables) จำนวน 6 ตัวแปร คือ ตัวแปรอายุ (age) มี 3 ระดับ เพศ (sex) มี 2 ระดับ ปริมาณน้ำดื่ม (drinkwater) มี 5 ระดับ ปริมาณน้ำประกอบอาหาร (cookwater) มี 5 ระดับ ปริมาณสารหนูในน้ำดื่ม (arswater) มี 2 ระดับ และปริมาณสารหนูในเล็บคน (arsnails) มี 2 ระดับจากข้อมูลจริงของการทดลองในมหาวิทยาลัย CMU (http://lib.stst.cmu.edu) การวิเคราะห์ข้อมูลอาศัย Fisher's exact test และตัวแบบล็อกลิเนียร์ โดยใช้โปรแกรม SAS version 9.1 ผลการวิจัยพบว่า ปัจจัยปริมาณสารหนูในน้ำดื่มมีผลกระทบต่อการปนเปื้อนหรือการสะสมของสารหนูในเล็บคนอย่างมีนัยสำคัญทาง สถิติที่ α = 0.05 ส่วนปัจจัยอายุ ตัวแปรเพศ ปริมาณน้ำดื่ม ปริมาณน้ำประกอบอาหาร ไม่มีผลกระทบต่อการปนเปื้อนของสารหนูในเล็บคนที่ α = 0.05 การวิเคราะห์ด้วยตัวแบบล็อกลิเนียร์ พบว่า ตัวแบบ $\log(\hat{m}_{ij}) = \hat{u} + \hat{u}_i^{arswater} + \hat{u}_j^{arsnails} + \hat{u}_{ij}^{arsnails*arswater}$ มีภาวะสารูปดีกับข้อมูล (P = 0.999) ดังนั้นอาจนำตัวแบบไปพยากรณ์ค่าคาดหมาย ($\hat{m}_{ij}$) จำแนกตามปริมาณสารหนูในเล็บคน ที่ j และปริมาณสารหนูในน้ำดื่ม ที่ i ($\hat{m}_{ij}$) ตลอดจนการนำไปพยากรณ์ค่าความน่าจะเป็นและความเสี่ยง (odds ratio) ที่สอดคล้องกับ $\hat{m}_{ij}$ ($\hat{P}_{ij}$) ต่อไป

[1]ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยศิลปากร วิทยาเขตพระราชวังสนามจันทร์ นครปฐม 73000

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

ABSTRACT

Arsenic contamination in human is possibly affecting one's body from several hazard factors. Public health and epidemiology works concern and monitor to this issue. One way to tackle this problem is to investigate several effecting factors that may associated with Arsenic cumulatively contamination in toenails. This research aims to study effective factorss on Arsenic contamination in toenails data obtained from website http://lib.stat.cmu.edu (8 March, 2010), on "Toenails Samples as an Indicator of Drinking Water Arsenic Exposure". The factors consist of 6 categorical variables: Age is categorized into 3 levels, similarly, for Sex is categorized into 2 levels, Drinkwater into 5 levels, Cookwater into 5 levels, Arswater into 2 levels, and Arsnails into 2 levels. Statistical analyses based on Fisher's exact test and loglinear modeling are performed using SAS version 9.1. The results under Fisher's exact test and those from log-linear modeling are concluded that for Fisher's tests, each Age, Sex, Cookwater, Drinkwater, is not associated with Arsnails under the level of significance at 0.05. However, it is found that the Arswater is associated with Arsnails significantly (p-value = 0.026). The analysis using loglinear models reveals that there is significantly associated between the Arsnails and Arswater at α = 0.05. The estimated loglinear model, $\log(\hat{m}_{ij}) = \hat{u} + \hat{u}_i^{arswater} + \hat{u}_j^{arsnails} + \hat{u}_{ij}^{arsnails*arswater}$ has adequate of fit (P = 0.999), of which the predicted values ($\hat{m}_{ij}$), the corresponding probabilities ($\hat{P}_{ij}$), and odds ratios can be obtained and further studied.

คำสำคัญ: สารหนู การปนเปื้อนสารหนูในคน การทดสอบของฟิชเชอร์ ตัวแบบล็อกลิเนียร์ การพยากรณ์

Keywords: Arsenic, Arsenic contamination in human, Fisher's exact test, Loglinear models, Forecasting

บทนำ

สารหนู (arsenic) เป็นธาตุกึ่งโลหะ เป็นสารที่มีลักษณะเป็นผงโลหะสีเทา มีมากเป็นอันดับที่ 20 ของธาตุที่พบมากบนโลก โดยจะพบในสิ่งที่มีชีวิตพืชและสัตว์ ตลอดจนพบในธรรมชาติ ได้แก่ ในพื้นดิน ทะเล มหาสมุทร และแหล่งน้ำต่าง ๆ โดยเป็นองค์ประกอบของดิน หิน และแร่ต่าง ๆ พบมากในสายแร่ทองแดง แมงกานีส ตะกั่ว ดีบุก เงินและทองคำ ในบริเวณแหล่งน้ำใต้ดินที่อยู่ใกล้สายแร่ดังกล่าว จึงมีการปนเปื้อนสารหนูสูง ซึ่งเป็นการปนเปื้อนโดยธรรมชาติ ส่วนการปนเปื้อนโดยการกระทำของมนุษย์ เช่น การจัดการเหมืองแร่ที่ไม่เหมาะสม การทำส่วนผสมหรือการปนเปื้อนในยาคน ยาสัตว์ และในอาหาร และอาจส่งผลกระทบต่อมาในอีกหลายด้าน มนุษย์จึงมีโอกาสได้รับสารหนูทั้งการสัมผัสและมีสารพิษสะสมมากขึ้นอยู่ตามกล้ามเนื้อกระดูก ผิวหนัง เล็บมือ เล็บเท้า สมอง ตับ และไต จนเป็นอันตราย (WHO, 1981) รวมทั้งการหายใจเอาละอองฝุ่นที่มีสารหนูหรือไอระเหยของสารหนู และการรับประทานอาหารหรือน้ำดื่มที่มีการปนเปื้อน โดยปกติร่างกายมนุษย์จะขจัดสารหนูที่ได้รับในปริมาณน้อย

ออกทางปัสสาวะภายในระยะเวลา 2 วัน แต่ถ้าหากได้รับสารหนูในปริมาณเพียง 130 มิลลิกรัม จะทำให้ลำไส้และตับถูกทำลาย อาเจียนมีสีเขียวและเหลือง ท้องเสียอย่างรุนแรง มึนเมา เพ้อ และถึงแก่ชีวิต หากได้รับปริมาณน้อยและมีระยะเวลานานเป็นเวลาหลายปี จะทำให้เกิดโรคพิษสารหนูเรื้อรังหรืออาร์ซินิโคซิส (arsenicosis) อาการจะเริ่มตั้งแต่ผิวหนังเกิดการระคายเคืองจนมีความด้านและหนา สีผิวเปลี่ยนเป็นสีดำเข้ม ฝ่ามือและฝ่าเท้าเป็นจุดสีดำใหญ่ จนกระทั่งกลายเป็นมะเร็งผิวหนัง มะเร็งปอด มะเร็งไต มะเร็งกระเพาะปัสสาวะ รวมทั้งมีผลต่อระบบหลอดเลือดหัวใจและระบบประสาทจนเสียชีวิตในที่สุด (อรุณศักดิ์, 2551) ในผู้ป่วยโรคมะเร็งผิวหนังอันเนื่องมาจากการปนเปื้อนสารหนู พบว่าระยะแฝงก่อนที่อาการของโรคปรากฏ (latent period) ของโรคมะเร็งผิวหนัง เท่ากับ 6 เดือน โดยแบ่งระยะดังนี้ ระยะ 0 ผลการศึกษาไม่พบรอยโรคมะเร็งผิวหนัง แต่พบว่ามีสารหนูในเล็บคนและเส้นผมในปริมาณค่อนข้างสูง ระยะ IA พบ pigment บริเวณฝ่ามือ ระยะ IB พบตุ่มน้ำเท่าหัวหมุด ระยะ II พบตุ่มน้ำจำนวนมาก ระยะ III พบ Bowenoid เปลี่ยนแปลง ระยะ IV เกิด Epithelioma และ Bowen's disease และ Squamous cell carcinoma 2 (สำนักงานพัฒนาระบบข้อมูลข่าวสารสุขภาพ, http://www.hiso.or.th)

วัตถุประสงค์ของงานวิจัยนี้ มีจุดมุ่งหมาย เพื่อศึกษาปัจจัยต่าง ๆ ที่มีผลต่อการปนเปื้อนของสารหนูในเล็บคน ซึ่งเป็นแหล่งข้อมูลสำคัญอันหนึ่งที่พบว่าอาจนำไปสู่อันตรายและมูลเหตุของโรคร้ายได้ และศึกษาความเกี่ยวพันของปัจจัยต่าง ๆ กับปริมาณของสารหนูในเล็บคน เพื่อการพยากรณ์ต่อไป

การวิเคราะห์ข้อมูลด้วย Fisher’s exact test (Fisher 1934, 1935a, 1935c; Yates, 1934) เป็นเทคนิคการวิเคราะห์ข้อมูลที่ใช้สำหรับตารางการจร (contingency table) ขนาดเล็ก หรือตัวอย่างขนาดเล็กดังข้อมูลการวิจัยครั้งนี้ ส่วนการวิเคราะห์ข้อมูลด้วยตัวแบบล็อกลิเนียร์ สมมติให้ตัวอย่างมัลติโนเมียลขนาด n ประกอบด้วย N = IJ เซลในตารางการจรสองทางขนาด (I x J) โดยมี $\{\mathrm{P}_{ij}\}$ แทนความน่าจะเป็นแบบมัลติโนเมียลหรือแทนการแจกแจงร่วมของตัวแปรเชิงกลุ่ม 2 ตัว และตัวแปรทั้งสองนี้จะเป็นอิสระต่อกันเมื่อ $\mathrm{P}_{ij} = \mathrm{P}_{i.}\mathrm{P}_{.j}$, i = 1,...,I และ j = 1,...,J นิพจน์ที่เกี่ยวข้องสำหรับความถี่ของค่าคาดหมาย (expected frequencies) ของ $\{\mathrm{m}_{ij} = \mathrm{nP}_{ij}\}$ คือ $\mathrm{m}_{ij} = \mathrm{nP}_{i.}\mathrm{P}_{.j}$ ในทุก ๆ i และ j และเราสามารถสร้างตัวแบบล็อกลิเนียร์ที่อาศัยเทอม $\{m_{ij}\}$ แทนเทอม $\{\mathrm{P}_{ij}\}$ โดยการประยุกต์กับลักษณะการแจกแจงแบบปัวส์ซองสำหรับความถี่ขนาด N เซล ซึ่งมีค่าคาดหมายเท่ากับ $\{m_{ij}\}$ ดังนั้น ตัวแบบล็อกลิเนียร์แบบง่าย ในเทอมของ m_{ij} จึงเริ่มต้นที่ตัวแบบที่เป็นอิสระต่อกัน (วีรานันท์, 2544, 2555) สรุปได้ดังนี้

จากเทอม $m_{ij} = n\mathrm{P}_{i.}\mathrm{P}_{.j}$ เราสามารถเขียนใหม่โดยใช้สเกลของล็อก (logarithmic scale) เพื่อทำให้ความอิสระต่อกันมีรูปแบบเป็นบวก (Agresti, 2002, 2007) ดังนี้

$$\log(m_{ij}) = \log(n) + \log(\mathrm{P}_{i.}) + \log(\mathrm{P}_{.j}) \tag{1}$$

ให้ตัวแบบทางด้านแถวแทนด้วย X และตัวแปรด้านสดมภ์แทนด้วย Y ดังนั้น (1) จะมีรูปแบบคล้ายตัวแบบการวิเคราะห์ความแปรปรวน (ANOVA model) คือ

$$\log(m_{ij}) = u + u_i^X + u_j^Y \tag{2}$$

เมื่อ
$$u_i^X = \log(P_{i.}) - \frac{\sum_h \log P_{h.}}{I}$$

$$u_j^Y = \log(P_{.j}) - \frac{\sum_h \log P_{.h}}{J}$$

$$u = \log(n) - \frac{\sum_h \log P_{h.}}{I} + \frac{\sum_h \log P_{.h}}{J}$$

และพารามิเตอร์ u_i^X และ u_j^Y จะสอดคล้องกับเงื่อนไข

$$\sum u_i^X = \sum u_j^Y = 0$$

สมการ (2) เรียกว่า ตัวแบบล็อกลิเนียร์ที่เป็นอิสระต่อกันในตารางการจรสองทาง (loglinear model of independence in a two – way contingency table)

สมมติว่ามีความไม่อิสระต่อกัน (dependence) ระหว่างตัวแปรต่าง ๆ และ $m_{ij} = nP_{ij} > 0$

ให้
$$\lambda_{ij} = \log(m_{ij})$$

เมื่อ
$$\lambda_{i.} = \frac{\sum_j \lambda_{ij}}{J},\ \lambda_{.j} = \frac{\sum_i \lambda_{ij}}{I},\ \lambda_{..} = \frac{\sum_i \sum_j \lambda_{ij}}{IJ} = u$$

โดยที่ u เป็นค่าเฉลี่ยทั้งหมด (grand mean) ของ $\log(m_{ij})$ นอกจากนี้

ให้
$$u_i^X = \lambda_{i.} - \lambda_{..},\ \ u_j^Y = \lambda_{.j} - \lambda_{..}$$

$$u_{ij}^{XY} = \lambda_{ij} - \lambda_{i.} - \lambda_{.j} + \lambda_{..} \tag{3}$$

ดังนั้น ตัวแบบล็อกลิเนียร์สำหรับตารางสองทาง คือ

$$\log(m_{ij}) = u + u_i^X + u_j^Y + u_{ij}^{XY} \tag{4}$$

สมการ (4) เรียกว่า ตัวแบบเต็ม (saturated model) ซึ่งเป็นตัวแบบที่สามารถอธิบายความเกี่ยวพันด้วยเซตของค่าคาดหมายของความถี่ที่เป็นบวกเซตใดเซตหนึ่งได้อย่างสมบูรณ์ และยังเป็นตัวแบบทั่วไปมากที่สุด (most general model) ของตารางการจรแบบสองทางนี้ สำหรับเทอมทางด้านขวาของสมการ (4) จะคล้ายคลึงกับสูตรของค่าเฉลี่ยสำหรับตารางวิเคราะห์ความแปรปรวนสองทาง (two- way ANOVA) ส่วนตัวแบบล็อกลิเนียร์สำหรับตารางหลายทาง เช่นสามทางสามารถทำได้ในทำนองเดียวกัน

วิธีการดำเนินการวิจัย

ข้อมูลเชิงกลุ่มที่ใช้ในการทำวิจัยในครั้งนี้ เป็นข้อมูลจริงล่าสุดจากเว็บไซต์ http://lib.stst.cmu.edu. (ค้นเมื่อ 8 มีนาคม 2553) เรื่อง Toenails Samples as an Indicator of Drinking Water Arsenic Exposure ปัจจัยที่สนใจศึกษาประกอบด้วย

1. อายุ (age) จำแนกเป็น 3 กลุ่ม คือ น้อยกว่า 40 ปี (กลุ่ม 1) ตั้งแต่ 40-59 ปี (กลุ่ม 2) และมากกว่า 59 ปี (กลุ่ม 3)

2. เพศ (sex) จำแนกเป็น 2 กลุ่ม คือ เพศชาย (กลุ่ม 1) และเพศหญิง (กลุ่ม 2)

3. ปริมาณน้ำที่ใช้สำหรับดื่ม (drinkwater) จำแนกเป็น 5 กลุ่ม คือ น้อยกว่า 1/4 ลิตรต่อครั้ง (กลุ่ม 1) 1/4 ลิตรต่อครั้ง (กลุ่ม 2) 1/2 ลิตรต่อครั้ง (กลุ่ม 3) 3/4 ลิตรต่อครั้ง (กลุ่ม 4) และมากกว่า 3/4 ลิตรต่อครั้ง (กลุ่ม 5)

4. ปริมาณน้ำที่ใช้สำหรับประกอบอาหาร (cookwater) จำแนกเป็น 5 กลุ่ม คือ น้อยกว่า 1/4 ลิตรต่อครั้ง (กลุ่ม 1) 1/4 ลิตรต่อครั้ง (กลุ่ม 2) 1/2 ลิตรต่อครั้ง (กลุ่ม 3) 3/4 ลิตรต่อครั้ง (กลุ่ม 4) และมากกว่า 3/4 ลิตรต่อครั้ง (กลุ่ม 5)

5. ปริมาณสารหนูในน้ำดื่ม (arswater) จำแนกเป็น 2 กลุ่มคือ น้อยกว่า 0.01 พีพีเอ็ม (กลุ่ม 1) และมากกว่า 0.01 พีพีเอ็ม (กลุ่ม 2)

6. ปริมาณสารหนูในเล็บคน (arsnails) จำแนกเป็น 2 กลุ่มคือ น้อยกว่า 0.6 พีพีเอ็ม (กลุ่ม 1) และมากกว่า 0.6 พีพีเอ็ม (กลุ่ม 2)

ตัวแบบที่นำมาวิเคราะห์ข้อมูลจำแนกประเภทและข้อมูลเชิงกลุ่มคือ ตัวแบบล็อกลิเนียร์ (loglinear model) การสร้างตัวแบบล็อกลิเนียร์ภายใต้ตัวแปรที่มีความเกี่ยวพันกับตัวแปรปริมาณสารหนูในเล็บคน อาศัยการเลือกตัวแปรเข้าไปในตัวแบบโดยจะรวมตัวแปรที่ให้ค่า p-value ของตัวสถิติ Fisher's exact test ที่น้อยกว่าระดับนัยสำคัญ α = 0.05 ด้วย ส่วนการทดสอบภาวะสารูปดี (goodness-of-fit tests) ของตัวแบบล็อกลิเนียร์ต่าง ๆ อาศัยค่าผลต่างของค่า Deviance (Agresti, 2002) ระหว่างสองตัวแบบ โดยใช้โปแกรม SAS เพื่อเปรียบเทียบว่าตัวแบบใดให้ความเหมาะสมกับข้อมูลมากที่สุดและมีความเกี่ยวพันระหว่างตัวแปรดีกว่ากัน ตัวแบบที่ได้จะประมาณค่าพารามิเตอร์และหาค่าความถี่ของค่าคาดหมายในแต่ละเซล

ผลการวิจัย

1. การศึกษาความเกี่ยวพันของตัวแปรต่าง ๆ โดยใช้สถิติ Fisher's exact test

ตารางที่ 1 ผลสรุปและค่า p-values จากการทดสอบความเกี่ยวพันด้วย Fisher's exact test

ตัวแปร	p-value	ผลสรุป
ปริมาณสารหนูในเล็บคนและอายุ	1.000	ไม่ปฏิเสธ H_0
ปริมาณสารหนูในเล็บคนและเพศ	0.257	ไม่ปฏิเสธ H_0
ปริมาณสารหนูในเล็บคนและปริมาณการใช้น้ำสำหรับดื่ม	0.384	ไม่ปฏิเสธ H_0
ปริมาณสารหนูในเล็บคนและปริมาณการใช้น้ำสำหรับประกอบอาหาร	1.000	ไม่ปฏิเสธ H_0
ปริมาณสารหนูในเล็บคนและตัวแปรปริมาณสารหนูในน้ำดื่ม	0.026	ปฏิเสธ H_0

จากตารางที่ 1 พบว่า ปัจจัยที่มีความเกี่ยวพันกับปริมาณสารหนูในเล็บคน คือ ปริมาณสารหนูในน้ำดื่มอย่างมีนัยสำคัญทางสถิติที่ 0.05 (p-value = 0.026) นอกนั้นไม่ปฏิเสธสมมติฐานของความเป็นอิสระกัน

2. การวิเคราะห์ปัจจัยที่มีผลต่อปริมาณของสารหนูในเล็บคน ทำการวิเคราะห์โดยตัวแบบล็อกลิเนียร์

ตารางที่ 2 ตัวประมาณพารามิเตอร์ของตัวแบบล็อกลิเนียร์ที่มีอิทิพลร่วมสองตัวแปร

Parameter	level	level	DF	Estimate	SE	Chi-Square	Pr > ChiSq
intercept			1	1.0986	0.5774	3.62	0.0571
arswater	1		1	-24.7918	0.5669	1912.19	<0.0001
arswater	2		0	0.0000	0.0000	.	.
arsnails	1		1	0.2877	0.7638	0.14	0.7064
arsnails	2		0	0.0000	0.0000	.	.
arswater*arsnails	1	1	1	26.0445	0.00001	2604450	0.0000
arswater*arsnails	1	2	0	0.0000	0.0000	.	.
arswater*arsnails	2	1	0	0.0000	0.0000	.	.
arswater*arsnails	2	2	0	1.0000	0.0000	.	.

จากตารางที่ 2 เป็นผลการวิเคราะห์ด้วยตัวแบบล็อกลิเนียร์ให้ค่าตัวประมาณพารามิเตอร์ของตัวแปรต่าง ๆ ที่ใช้กลุ่มหรือระดับสุดท้ายของแต่ละตัวแปรเป็นกลุ่มฐานเปรียบเทียบ (df=0) และอิทธิพลร่วมที่เกี่ยวกับกลุ่มฐานจึงมี df=0 ด้วย พบว่าปริมาณสารหนูในน้ำดื่มมีนัยสำคัญยิ่งทางสถิติ (p-value < 0.0001) และยืนยันความเกี่ยวพันต่อปริมาณสารหนูในเล็บคนอย่างมีนัยสำคัญยิ่งทางสถิติเช่นกัน (p-value = 0.0000)

ผลการประมาณค่าของพารามิเตอร์ของตัวแบบล็อกลิเนียร์ที่มีอิทธิพลร่วมสองตัวแปร (ตารางที่ 2) ด้วยรูปแบบตัวแบบล็อกลิเนียร์ที่มีอิทธิพลร่วมสองตัวแปร คือ

$$\log(m_{ij}) = u + u_i^{arswater} + u_j^{arsnails} + u_{ij}^{arsnails*arswater}$$

และได้ผลลัพธ์ ดังนี้

ตัวแบบล็อกลิเนียร์สำหรับเซล (1,1) คือ

$$\log(m_{11}) = u + u_1^{arswater} + u_1^{arsnails} + u_{11}^{arswater*arsnails}$$

$$\log(m_{11}) = 1.0986 - 24.7918 + 0.2877 + 26.0445 = 2.639$$

ตัวแบบล็อกลิเนียร์สำหรับเซล (1,2) คือ

$$\log(m_{12}) = u + u_1^{arswater} + u_2^{arsnails} + u_{12}^{arswater*arsnails}$$

$$\log(m_{12}) = 1.0986 - 24.7918 + 0 + 0 = -23.6932$$

ตัวแบบล็อกลิเนียร์สำหรับเซล (2,1) คือ

$$\log(m_{21}) = u + u_2^{arswater} + u_1^{arsnails} + u_{21}^{arswater*arsnails}$$

$$\log(m_{21}) = 1.0986 + 0 + 0.2877 + 0 = 1.3863$$

ตัวแบบล็อกลิเนียร์สำหรับเซล (2,2) คือ

$$\log(m_{22}) = u + u_2^{arswater} + u_2^{arsnails} + u_{22}^{arswater*arsnails}$$

$$\log(m_{22}) = 1.0986 + 0 + 0 + 1 = 2.0986$$

ผลลัพธ์นี้นำไปหาค่าคาดหมายภายใต้ระดับต่าง ๆ ของปริมาณสารหนูในเล็บคนและปริมาณสารหนูในน้ำดื่มต่อไป

ทำนองเดียวกันกับตัวประมาณพารามิเตอร์ของตัวแบบล็อกลิเนียร์ที่มีอิทธิพลร่วมสองตัวแปร จะสามารถประมาณค่าของพารามิเตอร์ของตัวแบบล็อกลิเนียร์ที่มีอิทธิพลร่วมสามตัวแปร ด้วยรูปแบบตัวแบบล็อกลิเนียร์ที่มีอิทธิพลร่วมสามตัวแปร คือ

$$\log(m_{ijk}) = u + u_i^{cookuse} + u_j^{arswater} + u_k^{arsnails} + u_{ij}^{cookuse*arswater} + u_{ik}^{cookuse*aesnails} + u_{jk}^{arswater*arsnails} + u_{ijk}^{cookuse*arswater*arsnails}$$

และได้ผลลัพธ์ ดังนี้

ตัวแบบล็อกลิเนียร์สำหรับเซล (1,1,1) คือ

$$\log(m_{111}) = u + u_1^{cookuse} + u_1^{arswater} + u_1^{arsnails} + u_{11}^{cookuse*arswater} + u_{11}^{cookuse*arsnails} + u_{11}^{arswater*arsnails} + u_{111}^{cookuse*arswater*arsnails}$$

$$\log(m_{111}) = 1.0986 - 24.7918 - 24.7917 + 24.7917 + 0.2877 + 25.9704 - 25.9704 = -23.4055$$

ตัวแบบล็อกลิเนียร์สำหรับเซล (1,1,2) คือ

$$\log(m_{112}) = u + u_1^{cookuse} + u_1^{arswater} + u_2^{arsnails} + u_{11}^{cookuse*arswater} + u_{12}^{cookuse*arsnails} + u_{12}^{arswater*arsnails} + u_{112}^{cookuse*arswater*arsnails}$$

$$\log(m_{112}) = 1.0986 - 24.7918 - 24.7917 + 0 + 24.7917 + 0 + 0 + 0 = -23.6932$$

ตัวแบบล็อกลิเนียร์สำหรับเซล (5,2,1) คือ

$$\log(m_{521}) = u + u_5^{cookuse} + u_2^{arswater} + u_1^{arsnails} + u_{52}^{cookuse*arswater} + u_{51}^{cookuse*arsnails} + u_{21}^{arswater*arsnails} + u_{521}^{cookuse*arswater*arsnails}$$

$$\log(m_{521}) = 1.0986 + 0 + 0 + 0.2877 + 0 + 0 + 0 + 0 = 1.3863$$

ตัวแบบล็อกลิเนียร์สำหรับเซล (5,2,2) คือ

$$\log(m_{522}) = u + u_5^{cookuse} + u_2^{arswater} + u_2^{arsnails} + u_{52}^{cookuse*arswater} + u_{52}^{cookuse*arsnails} + u_{22}^{arswater*arsnails} + u_{522}^{cookuse*arswater*arsnails}$$

$$\log(m_{522}) = 1.0986 + 0 + 0 + 0 + 0 + 0 + 0 + 0 = 1.0986$$

ตารางที่ 3 การประเมินภาวะสารูปดี (goodness of fit) ของตัวแบบที่มีอิทธิพลร่วมสองตัวแปรและสามตัวแปร

Criterion	DF		Value	
ตัวสถิติ	**ตัวแบบอิทธิพลร่วมสองตัวแปร**	**ตัวแบบอิทธิพลร่วมสามตัวแปร**	**ตัวแบบอิทธิพลร่วมสองตัวแปร**	**ตัวแบบอิทธิพลร่วมสามตัวแปร**
Deviance	1	13	7.6642	8.4999
Scaled Deviance	1	13	7.6642	8.4999
Pearson Chi-Square	1	13	7.0000	7.7875
Scaled Pearson X2	1	13	7.0000	7.7875

จากตารางที่ 3 ค่า Deviance ของตัวแบบที่มีอิทธิพลร่วมสองตัวแปรเท่ากับ 7.6642 และของตัวแบบที่มีอิทธิพลร่วมสามตัวแปรเท่ากับ 8.4999

จากค่า Deviance ของทั้งสองตัวแบบเมื่อนำมาคำนวณผลต่างของค่า Deviance จะได้ผลต่างของค่า Deviance (ΔDev) = ค่า Deviance ของตัวแบบที่มีอิทธิพลร่วมสามตัวแปร – ค่า Devianceของตัวแบบที่มีอิทธิพลร่วมสองตัวแปร = 8.4999-76642 = 0.8357

การทดสอบความเกี่ยวพันระหว่างปริมาณสารหนูในเล็บคนและปริมาณสารหนูในน้ำดื่ม ว่ามีความเกี่ยวพันกันหรือไม่ ทำโดยทดสอบสมติฐานสำหรับตัวแบบที่มีอิทธิพลร่วมสองตัวแปร (H_o) กับตัวแบบที่มีอิทธิพลร่วมสามตัวแปร (H_1) ดังนี้

$$\mathrm{H}_0 : \log(m_{ij}) = u + u_i^{arswater} + u_j^{arsnails} + u_{ij}^{arsnails*arswater}$$

$$\mathrm{H}_1 : \log(m_{ijk}) = u + u_i^{cookuse} + u_j^{arswater} + u_k^{arsnails} + u_{ij}^{cookuse*arswater} + u_{ik}^{cookuse*aesnails}$$
$$+ u_{jk}^{arswater*arsnails} + u_{ijk}^{cookuse*arswater*arsnails}$$

ผลลัพธ์จาก $\chi^2_{12,0.05}$ = 21.0261 ซึ่งมากกว่า ผลต่างของค่า Deviance = 0.8357 จึงไม่ปฏิเสธ H_0 นั่นคือตัวแบบล็อกลิเนียร์สองตัวแปรคือ $\mathrm{H}_0 : \log(m_{ij}) = u + u_i^{arswater} + u_j^{arsnails} + u_{ij}^{arsnails*arswater}$ มีความเหมาะสมกับข้อมูลอย่างมีนัยสำคัญทางสถิติที่ 0.05 และสรุปว่า ปริมาณสารหนูในน้ำดื่มอาจส่งผลกระทบต่อการพบปริมาณสารหนูในเล็บคน

เนื่องจากตัวแบบที่เหมาะสมประกอบด้วยตัวแปรเชิงกลุ่ม 2 ตัว คือ ตัวแปรปริมาณสารหนูในเล็บคน และตัวแปรปริมาณสารหนูในน้ำดื่ม ซึ่งต่างก็มี 2 ระดับ ดังนั้นตารางการจรจึงมีขนาด (2×2) และตัวแบบล็อกลิเนียร์คือ

$$\log(m_{ij}) = u + u_i^{arswater} + u_j^{arsnails} + u_{ij}^{arsnails*arswater}, \text{ i, j = 1, 2}$$

โดยมีค่าคาดหมาย เช่น $\log(m_{11}) = u + u_1^{arswater} + u_1^{arsnails} + u_{11}^{arswater*arsnails}$

การประมาณพารามิเตอร์กำหนดให้ ระดับที่ 2 ของแต่ละตัวแปร เป็น reference หรือ baseline ซึ่งจะเห็นว่าเทอมของ u_{ij} ที่เกี่ยวข้องกับระดับที่ 2 เป็น 0 ภายใต้สมมติฐาน $\sum u_i = \sum u_j = \sum u_{ij} = 0$

สรุปการคำนวณค่าคาดหมายหรือค่าประมาณความถี่ของค่าคาดหมาย $(\hat{m}_{ij})$ ซึ่งคำนวณจากผลลัพธ์ในตารางที่ 2 ดังนี้

ตัวแบบล็อกลิเนียร์สำหรับเซล (1,1) คือ

$$\log(m_{11}) = 2.639\ ,\ \hat{m}_{11} = \exp(2.639) = 13.9992$$

การประมาณพารามิเตอร์ด้วยตัวแบบล็อกลิเนียร์ จะได้ความถี่ของค่าคาดหมายของเซล (1,1) เท่ากับ 13.99

ตัวแบบล็อกลิเนียร์สำหรับเซล (1,2) คือ

$$\log(m_{12}) = -23.6932\ ,\ \hat{m}_{12} = \exp(-23.6932) = 5.1307\times10^{-11}$$

การประมาณพารามิเตอร์ด้วยตัวแบบล็อกลิเนียร์ จะได้ความถี่ของค่าคาดหมายของเซล (1,2) เท่ากับ 5.13×10^{-11}
ตัวแบบล็อกลิเนียร์สำหรับเซล (2,1) คือ

$$\log(m_{21}) = 1.3863\ ,\ \hat{m}_{21} = \exp(1.3863) = 4.0000$$

การประมาณพารามิเตอร์ด้วยตัวแบบล็อกลิเนียร์ จะได้ความถี่ของค่าคาดหมายของเซล (2,1) เท่ากับ 4.0000
ตัวแบบล็อกลิเนียร์สำหรับเซล (2,2) คือ

$$\log(m_{22}) = 1.0986\ ,\ \hat{m}_{22} = \exp(2.0986) = 8.1547$$

จากการประมาณพารามิเตอร์ด้วยตัวแบบล็อกลิเนียร์ จะได้ความถี่ของค่าคาดหมายของเซล (2,2) เท่ากับ 8.1547

ดังนั้นค่าประมาณความถี่ค่าคาดหมายและค่าสังเกตของแต่ละเซล ซึ่งเป็นประโยชน์สำคัญของตัวแบบล็อกลิเนียร์ในการพยากรณ์เซลข้อมูล (วิสูตร และวีรานันท์, 2553) และในการพยากรณ์ค่าความน่าจะเป็นจากผลลัพธ์ของตัวแบบล็อกลิเนียร์ต่อไป จะแสดงไว้ร่วมกันในตารางที่ 4-5

ตารางที่ 4 ค่าสังเกตจากข้อมูลและค่าประมาณความถี่ของค่าคาดหมายของแต่ละเซลในตารางการจร (2×2)

ปัจจัย		ระดับปริมาณสารหนูในเล็บคน (พีพีเอ็ม)	
		น้อยกว่า 0.6	มากกว่า 0.6
ตัวแปรปริมาณสารหนูในน้ำดื่ม (พีพีเอ็ม)	น้อยกว่า 0.01	$O_{11} = 14, \hat{m}_{11} = 13.9992$	$O_{12} = 0, \hat{m}_{12} = 5.1307\times10^{-11}$
	มากกว่า 0.01	$O_{21} = 4, \hat{m}_{21} = 4.0000$	$O_{22} = 3, \hat{m}_{22} = 8.1547$

ตารางที่ 5 การพยากรณ์ความน่าจะเป็นของเซลจากตัวแบบล็อกลิเนียร์

ปัจจัย		ระดับปริมาณสารหนูในเล็บคน (พีพีเอ็ม)	
		น้อยกว่า 0.6	มากกว่า 0.6
ระดับปริมาณสารหนูในน้ำ ดื่ม (พีพีเอ็ม)	น้อยกว่า 0.01	$\hat{P}_{11} = 0.5353$	$\hat{P}_{12} = 1.9615\times10^{-12}$
	มากกว่า 0.01	$\hat{P}_{21} = 0.1529$	$\hat{P}_{22} = 0.3118$

จากตารางที่ 5 ความน่าจะเป็นของเซลจากตัวแบบล็อกลิเนียร์สองตัวแปรสามารถคำนวณจากสูตรที่ใช้ค่าคาดหมายของเซลในตารางที่ 4 ดังนี้

$$P_{ij} = m_{ij} / \sum m_{ij}, i = 1,2, j = 1,2$$

พบว่า การพยากรณ์ความน่าจะเป็นของตัวแปรสารหนูในน้ำและตัวแปรสารหนูในเล็บคนที่สูงสุด คือ 0.666 รองลงมาคือ 0.190 ตามลำดับ

และพบว่า อัตราส่วนออดส์ (odds ratio) ของเซลจากตัวแบบล็อกลิเนียร์ เท่ากับ

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

ซึ่งสามารถประมาณด้วย

$$\hat{\theta} = \frac{\hat{m}_{11}\hat{m}_{22}}{\hat{m}_{12}\hat{m}_{21}} = \frac{13.9992 \times 8.1547}{4.0000 \times 5.1307 \times 10^{-11}} = 5.5633 \times 10^{11}$$

หรือส่วนกลับคือ $\frac{1}{5.5633 \times 10^{11}} = 1.7975 \times 10^{-12}$

หมายความว่า ปริมารสารหนูในน้ำดื่มที่น้อยกว่า 0.01 พีพีเอ็มมีโอกาสเสี่ยงต่อการมีปริมารสารหนูในเล็บคนที่น้อยกว่า 0.6 พีพีเอ็ม เป็น 1.7975×10^{-12} เท่าของปริมาณสารหนูในน้ำดื่มที่มากกว่า 0.01 พีพีเอ็ม

กล่าวอีกนัยหนึ่งปริมาณสารหนูในน้ำดื่มที่มากกว่า 0.01 พีพีเอ็ม มีโอกาสเสี่ยงมากต่อการมีสารหนูในเล็บคนมากเกิน 0.06 พีพีเอ็มถึงประมาณ 5.5633×10^{11} เท่าของกรณีปริมารสารหนูในน้ำที่น้อยกว่า 0.01 พีพีเอ็ม

วิจารณ์ผลการวิจัย

ผลการศึกษาความเกี่ยวพันของปัจจัยต่าง ๆ โดยใช้สถิติ Fisher's Exact Test และลักษณะทั่วไปของปัจจัยต่าง ๆ กับปริมาณสารหนูในเล็บคน พบว่าปริมาณสารหนูในเล็บคนและปริมาณสารหนูในน้ำดื่มมีความเกี่ยวพันกัน ที่ระดับนัยสำคัญ 0.05 ส่วนปริมาณสารหนูในเล็บคนและปัจจัย อื่น ๆ เช่น อายุ เพศ ปริมาณน้ำที่ใช้สำหรับดื่ม และปริมาณน้ำที่ใช้สำหรับประกอบอาหาร ไม่พบความเกี่ยวพันกันเชิงสถิติ ที่ระดับนัยสำคัญ 0.05

ผลการสร้างตัวแบบล็อกลิเนียร์ ภายใต้ปัจจัยที่มีผลต่อปริมาณของสารหนูในเล็บคน และปริมารสารหนูในน้ำ ดื่มได้ตัวแบบล็อกลิเนียร์ ดังนี้

$$\log(m_{ij}) = u + u_i^{arswater} + u_j^{arsnails} + u_{ij}^{arsnails*arswater}$$

เมื่อ $u = 1.0986$, $u_i^{arswater} = -24.7918$, $u_j^{arsnails} = 0.2877$, $u_{ij}^{arsnails*arswater} = 26.0445$

การพยากรณ์ค่าคาดหมายด้วยตัวแบบล็อกลิเนียร์สำหรับเซล (i, j), i = 1, 2 และ j = 1, 2 พบว่าค่าประมาณความถี่ของค่าคาดหมายและค่าสังเกตของเซล (1,1) เท่ากับ 13.9992 และ 14 ตามลำดับ ทำนองเดียวกัน ค่าประมาณความถี่ของค่าคาดหมายและค่าสังเกตของเซล (1,2) เท่ากับ 5.1307×10^{-11} และ 0 ค่าประมาณความถี่ของค่าคาดหมายและค่าสังเกตของเซล (2,1) เท่ากับ 4.0000 และค่าประมาณความถี่ของค่าคาดหมายและค่าสังเกตของเซล (2,2) เท่ากับ 8.1547 และ 3 ตามลำดับ ซึ่งสามารถนำไปวิเคราะห์ข้อมูลเพิ่มเติม เช่น ความน่าจะเป็นต่าง ๆ และค่า odds แบบต่าง ๆ ต่อไป โดยสรุปพบว่า ปัจจัยที่มีผลต่อปริมาณของสารหนูในเล็บคน คือ ปริมาณของสารหนูในน้ำดื่ม นั่นคือ ถ้ามีปริมาณสารหนูปนเปื้อนในน้ำดื่มากจะมีโอกาสพบปริมาณการปนเปื้อนของสารหนูในเล็บคนมาก ด้วยอย่างมีนัยสำคัญทางสถิติที่ 0.05

อย่างไรก็ตาม งานวิจัยครั้งต่อไปถ้าสามารถศึกษาด้วยตัวอย่างที่เพิ่มมากขึ้น น่าจะได้ผลสรุปที่ยืนยันชัดเจนยิ่งขึ้นต่อไป เนื่องจากปัจจัยที่อาจมีผลกระทบต่อปริมาณของสารหนูในเล็บคน คือ ปริมาณของสารหนูในน้ำดื่มอาจต้องการสารสนเทศเพิ่มเติมจากการวิจัยเพิ่มเติมต่อไปว่า ปริมาณของสารหนูในน้ำดื่มที่พบนี้ จะส่งผล

กระทบเชิงสถิติ ต่อปริมาณน้ำดื่มและ/หรือปริมาณน้ำที่ใช้ประกอบอาหารมากขึ้นหรือไม่ (http://www.WHO.int/wer/en)

ถึงแม้ว่าการเพิ่มขนาดตัวอย่าง โดยเฉพาะรวมถึงการเพิ่มตัวแปรของการวิจัยให้มากขึ้นด้วย อาจทำให้มีค่าสังเกตในบางเซลเป็น 0 ได้ แต่กรณีเช่นนี้ยังสามารถอาศัยวิธีการของตัวแบบล็อกลิเนียร์มาวิเคราะห์ความเกี่ยวพันแบบต่าง ๆ เพื่อศึกษาปัจจัยที่อาจส่งผลกระทบ และเพื่อการหาค่าประมาณของความถี่คาดหมายจากตัวแบบอันนำไปสู่การประเมินผลลัพธ์ที่สนใจ รวมทั้งการประมาณความน่าจะเป็นของเซลได้ (Agresti, 2002, 2007)

สรุปผลการวิจัย

จากผลของการวิจัยในครั้งนี้พบว่า ปริมาณสารหนูในน้ำดื่มอาจเป็นปัจจัยที่ส่งผลต่อปริมาณการพบสารหนูในเล็บคนอย่างมีนัยสำคัญทางสถิติที่ระดับนัยสำคัญ 0.05 แต่อาจยังมีปัจจัยอื่น ๆ ที่เกี่ยวข้องอื่น ๆ เช่น ชนิดของแหล่งน้ำ คือ เป็นแหล่งน้ำใต้ดินหรือผิวดินหรือการประปา ซึ่งเป็นปัจจัยอีกปัจจัยหนึ่งที่อาจส่งผลต่อการพบปริมาณสารหนูในเล็บคน นอกจากนี้การจำแนกกลุ่มข้อมูลออกเป็นกลุ่ม ๆ ตามเอกสารข้อมูลที่นำมาวิเคราะห์เช่นเรื่อง Toenail Samples as an Indicator of Drinking Water Arsenic Exposure และขององค์การอนามัยโลกหรือของกรมวิทยาศาสตร์การแพทย์ และกระทรวงสาธารณสุข (http://www2.dmsc.moph.go.th /web/DMSC.LIB/index.html/) นั้น ในกรณีซึ่ง ถ้ามีการศึกษาตัวแปรเพิ่มเติมด้วย อาจมีการจำแนกกลุ่มของตัวแปรตามเกณฑ์อื่น ๆ ของผู้เชี่ยวชาญได้

เอกสารอ้างอิง

สำนักงานพัฒนาระบบข้อมูลข่าวสารสุขภาพ http://www.hiso.or.th (ค้นเมื่อ 18 กันยายน 2555)

วิสูตร สาลี และวีรานันท์ พงศาภักดี. (2553). การพยากรณ์ระดับความแห้งแล้งด้วยตัวแบบล็อกลิเนียร์ภายใต้ SPI ในภาคตะวันออกเฉียงเหนือของประเทศไทย. การประชุมวิชาการสถิติและสถิติประยุกต์ครั้งที่ 11 ประจำปี 2553, เชียงใหม่. หน้า 51-58.

วีรานันท์ พงศาภักดี. (2544). การวิเคราะห์ข้อมูลเชิงกลุ่ม: ทฤษฎีและการประยุกต์. โรงพิมพ์มหาวิทยาลัยศิลปากร

วีรานันท์ พงศาภักดี. (2555). การวิเคราะห์ข้อมูลจำแนกประเภท: ทฤษฎีและการประยุกต์ ด้วย GLIM, SPSS, SAS และ MTB โรงพิมพ์มหาวิทยาลัยศิลปากร

อรุณศักดิ์ โสภณธรรมภาณ. (2551). อันตรายจากการปนเปื้อนของสารหนูในแหล่งน้ำ. นักวิชาการ สำนักวิชาการสถาบันวิจัยจุฬาภรณ์

Agresti, A. (2002). Categorical Data Analysis. 2nd edition. New York: John Willey & Sons.

Agresti, A. (2007). Introduction to Categorical Data Analysis. 2nd edition. New York: John Willey & Sons.

Fisher, R.A. (1934). Statistical Methods for Research Workers. 14th edition. Edinburgh: Oliver & Boyd.

Fisher, R.A. (1935a). The Design of Experiments. Edinburgh: Oliver & Boyd.

Fisher, R.A. (1935c). The logic of inductive inference. J. Roy. Statist. Soc. 98: 39-82.

Yates, F. (1934). Contingency tables involving small numbers and the χ^2 test. J. Roy. Statist. Soc. Suppl. 1: 217-235.

WHO (1981). Arsenic Environment health criteria 18. World Health Organization.

http://lib.stat.cmu.edu/datasets/Arsenic (ค้นเมื่อ 8-10 มีนาคม 2553)
http://www.WHO.int/wer/en (ค้นเมื่อ 8-10 มีนาคม 2553)
http://www2.dmsc.moph.go.th/web/DMSC.LIB/index.html/ (ค้นเมื่อ 8-10 มีนาคม 2553)

❑❑❑❑❑