

ปัจจัยเสี่ยงที่มีความสัมพันธ์ต่อการเกิดโรคมะเร็ง

Risk factors for cancer

พิมพจุฑา วราทรัพย์^{1*} จิราวัลย์ จิตรลเวช² และ วิชิต หล่อจรัสชุมพุก³

Pimjuta Varasap¹ Jirawan Jitthavech² and Vichit Lorchorchoonkul³

¹สาขาวิชาสถิติ คณะสถิติประยุกต์ สถาบันบัณฑิตพัฒนบริหารศาสตร์

²ศาสตราจารย์ สาขาวิชาสถิติ คณะสถิติประยุกต์ สถาบันบัณฑิตพัฒนบริหารศาสตร์

³รองศาสตราจารย์ สาขาวิชาสถิติ คณะสถิติประยุกต์ สถาบันบัณฑิตพัฒนบริหารศาสตร์

วันที่ส่ง : 28 กันยายน 2560 วันที่แก้ไข : 30 พฤศจิกายน 2560 วันที่ตอบรับ : 30 พฤศจิกายน 2561

บทคัดย่อ

การวิจัยครั้งนี้ เป็นการประยุกต์ใช้การวิเคราะห์การถดถอยลอจิสติกเพื่อศึกษาปัจจัยเสี่ยงที่มีความสัมพันธ์ต่อการเกิดโรคมะเร็ง ข้อมูลที่ใช้ในงานวิจัยครั้งนี้เป็นข้อมูลทุติยภูมิของโครงการสำรวจพฤติกรรมเสี่ยงโรคไม่ติดต่อและการบาดเจ็บ พ.ศ. 2558 สำนักโรคไม่ติดต่อ กรมควบคุมโรค กระทรวงสาธารณสุข และวิเคราะห์ข้อมูลโดยใช้โปรแกรมสำเร็จรูป SAS (Statistical Analysis System) เวอร์ชัน 9.4 และ SAS Enterprise Miner Workstation เวอร์ชัน 14.2

เนื่องจากข้อมูลที่นำมาใช้ในงานวิจัยเป็นข้อมูลที่ไม่สมดุล คือมีจำนวนผู้ป่วยมะเร็งและผู้ที่ไม่เป็นมะเร็งแตกต่างกันมาก ผู้วิจัยจึงทำการปรับข้อมูลให้สมดุลด้วยวิธีการสุ่มลดจำนวนผู้ที่ไม่เป็นมะเร็ง แล้วพัฒนาตัวแบบทำนายความน่าจะเป็นของการเกิดโรคมะเร็งจากการเปรียบเทียบประสิทธิภาพของตัวแบบ พบว่าตัวแบบที่ดีที่สุดคือตัวแบบที่สร้างจากข้อมูลที่สุ่มลดจำนวนผู้ที่ไม่เป็นมะเร็งให้มีจำนวนผู้ป่วยมะเร็ง : ผู้ที่ไม่เป็นมะเร็ง เท่ากับ 1 : 3 ผลการวิจัยสรุปได้ว่า สมการที่สร้างขึ้นมีความเหมาะสมกับข้อมูลเชิงประจักษ์ ด้วยค่าไคกำลังสองเท่ากับ 0.2464 และค่าพีเท่ากับ 0.8841 และมีร้อยละความถูกต้อง 76.80 ร้อยละความไว 13.10 และร้อยละความเฉพาะ 98.00 เมื่อกำหนดความน่าจะเป็นในการแบ่งกลุ่มเป็น 0.5 เมื่อพิจารณาอัตราส่วน Odds ของตัวแปรอิสระแต่ละตัวพบว่า เพศหญิงมีความเสี่ยงที่จะเป็นมะเร็งมากกว่าเพศชาย 7.3440 เท่า คนสูบบุหรี่มีความเสี่ยงที่จะเป็นมะเร็งมากกว่าคนที่ไม่สูบบุหรี่ 6.1540 เท่า และคนที่มียุมากกว่าหรือเท่ากับ 45 ปีมีความเสี่ยงที่จะเป็นมะเร็งมากกว่าคนที่มียุน้อยกว่า 45 ปี 4.6360 เท่า

คำสำคัญ: ปัจจัยเสี่ยงของโรคมะเร็ง การวิเคราะห์การถดถอยลอจิสติก ความถูกต้องเหมาะสมของตัวแบบอัตราส่วน Odds ความถูกต้อง ความไว ความเฉพาะ

*E-mail : pimjuta.va@gmail.com

Abstract

The objective of the research is to apply the logistic regression analysis to investigate factors related to cancer. The data use in this research is the secondary data from Thai Behavioral Risk Factor Surveillance by the Questionnaire in BE. 2558, Bureau of Non-communicable Disease, Department of Disease Control, Ministry of Public Health. The software SAS version 9.4 and SAS Enterprise Miner Workstation version 14.2 are used to analyze and develop the logistic regression model for prediction the probability of being a cancer-patient.

Since the numbers of cancer patients and non-cancer patients are very different the problem of imbalanced data is to be corrected before developing the predictive model by under sampling the non-cancer patients. From the comparison of the results from several models based on data sets with different ratios of cancer patients over non-cancer patients, it is found that the predictive model based on the data set of with ratio of cancer patients over non-cancer patients equal of 1/3 gives the best result. The independent variables for predicting the probability of developing a cancer are gender, smoking and age.

The developed model is fit to the empirical data with chi-squared of 0.2464, p-value = 0.8841 at df = 3. When the cutoff probability equals 0.5, the specificity, sensitivity and accuracy of the model are 98.00 percent, 13.10 percent and 76.80 percent respectively. The odds ratio of developing a cancer given being female to being male is 7.3440, of developing a cancer given smoking to non-smoking is 6.1540 and of developing a cancer given age 45 year and over to less than 45 year is 4.6360.

Keywords: risk factors for cancer, Logistic regression analysis, Goodness of fit, Odds ratio, accuracy, sensitivity, specificity

1. บทนำ

โรคมะเร็งเป็นสาเหตุของการเสียชีวิตอันดับต้นๆ ของผู้คนทั่วโลก และมีแนวโน้มว่าจำนวนผู้เสียชีวิตจากโรคมะเร็งจะเพิ่มสูงขึ้นในทุกๆ ปี สำหรับประเทศไทยนั้นโรคมะเร็งก็ถือว่าเป็นสาเหตุของการเสียชีวิตอันดับ 1 ของคนไทย ซึ่งข้อมูลล่าสุดจากกระทรวงสาธารณสุข พบว่าคนไทยเสียชีวิตจากโรคมะเร็งประมาณ 60,000 คนต่อปี หรือเฉลี่ยชั่วโมงละเกือบ 7 ราย โดยมีสาเหตุการเสียชีวิตจากโรคมะเร็ง 5 อันดับแรก ได้แก่ มะเร็งตับและท่อน้ำดี มะเร็งปอด มะเร็งเต้านม มะเร็งลำไส้ใหญ่และทวารหนัก และมะเร็งปากมดลูก โรคมะเร็งเป็นภัยร้ายที่คร่าชีวิตผู้คนไปเป็นจำนวนมาก การป้องกันความทุกข์ทรมานและการตายจากโรคมะเร็งนั้น จึงถือเป็นเรื่องใหญ่สำหรับทุกประเทศ ซึ่งในปัจจุบันสาเหตุของการเกิด

โรคมะเร็งยังไม่ชัดเจน แต่มีงานวิจัยที่ทำให้เชื่อได้ว่าสาเหตุของการเกิดโรคมะเร็งน่าจะมาจากหลาย ๆ ปัจจัยร่วมกัน โดยปัจจัยเสี่ยงสำคัญของโรคมะเร็ง ได้แก่ การสูบบุหรี่ การดื่มสุราขาดสารอาหารการรับประทานผักและผลไม้สดน้อยรับประทานไขมันสูงต่อเนื่องเป็นประจำ สูงอายุ และการสูดดมสารพิษบางชนิดเรื้อรัง เช่น สารพิษในควันบุหรี่ เป็นต้น[1, 2]การป่วยเป็นโรคมะเร็งมีผลกระทบต่อสภาพร่างกายและจิตใจของผู้ป่วย ครอบครัว คนรอบข้างของผู้ป่วย และหน่วยงานต่าง ๆ ที่มีผู้ป่วยเป็นมะเร็งอาจต้องขาดแคลนบุคลากรที่สำคัญของหน่วยงานไป การศึกษาค้นคว้าและรวบรวมปัจจัยที่ก่อให้เกิดโรคมะเร็ง ทำให้ทราบปัจจัยที่มีความสัมพันธ์ต่อการเกิดโรคมะเร็ง และนำผลลัพธ์ที่ได้จากการศึกษาไปใช้อธิบายว่าปัจจัยใดบ้างที่ทำให้เกิดโรคมะเร็ง ตลอดจนเป็นแนวทางการลดความเสี่ยงในการเกิดโรคมะเร็งในอนาคตได้

2. วิธีการวิจัย

ในการศึกษารังนี้ผู้วิจัยใช้ข้อมูลทุติยภูมิของโครงการสำรวจพฤติกรรมเสี่ยงโรคไม่ติดต่อและการบาดเจ็บ พ.ศ. 2558 ของสำนักโรคไม่ติดต่อ กรมควบคุมโรค กระทรวงสาธารณสุข ซึ่งตัวแปรการเกิดโรคมะเร็งไม่ได้มีการเก็บข้อมูลชนิดของมะเร็งจึงไม่สามารถศึกษาปัจจัยเสี่ยงแยกตามชนิดของมะเร็งได้ การวิเคราะห์ข้อมูลใช้สถิติเชิงพรรณนาในการอธิบายข้อมูลเบื้องต้นของตัวอย่าง และประยุกต์ใช้การวิเคราะห์การถดถอยลอจิสติก (Logistic regression) เพื่อศึกษาความสัมพันธ์ของตัวแปรอิสระที่มีต่อตัวแปรตามคือการเกิดมะเร็งโดยอาศัยสมการลอจิสติกที่สร้างขึ้นจากชุดตัวแปรอิสระ[3] ซึ่งมีขั้นตอนในการดำเนินงานวิจัยดังนี้

2.1 การเตรียมข้อมูล

จากการทบทวนวรรณกรรมและงานวิจัยที่เกี่ยวข้องกับการเกิดโรคมะเร็ง พบว่ามีปัจจัยเสี่ยงที่อาจเพิ่มโอกาสในการเป็นโรคมะเร็งอยู่หลายปัจจัย ได้แก่

บุหรี่ อาจถือได้ว่าบุหรี่เป็นสาเหตุสำคัญที่ทำให้เกิดมะเร็งปอด เนื่องจากสารพิษในบุหรี่สามารถทำลายเซลล์ปอด และทำให้เกิดความผิดปกติของเซลล์ผู้ที่สูบบุหรี่มีความเสี่ยงที่จะเป็นมะเร็งปอดมากกว่าผู้ที่ไม่สูบบุหรี่โดยความเสี่ยงจะเพิ่มขึ้นตามจำนวนมวนและจำนวนปีที่สูบบุหรี่[1, 4]

การรับประทานผักและผลไม้สดน้อยมีหลักฐานทางวิทยาศาสตร์ที่บ่งชี้ว่า การรับประทานผักผลไม้สดอย่างเพียงพอจะช่วยลดความเสี่ยงของมะเร็งหลอดอาหาร, มะเร็งกระเพาะอาหาร, มะเร็งตับอ่อน และมะเร็งลำไส้ใหญ่เนื่องจากผักและผลไม้สดมีกากใยสูงและมีคุณสมบัติต้านอนุมูลอิสระ เมื่อใยอาหารเคลื่อนเข้าสู่ลำไส้ใหญ่ย่อมมีผลทำให้อุจจาระอ่อนตัวช่วยให้ขับถ่ายของเสียออกนอกร่างกายง่ายขึ้น [5,6]

อายุ ความเสี่ยงของการเกิดโรคมะเร็งจะเพิ่มขึ้นตามอายุ โดยทั่วไปความเสี่ยงจะเพิ่มขึ้นหลังอายุ 45ปี แต่ก็สามารถพบได้ในคนที่อายุน้อยกว่า 45ปี เมื่ออายุเพิ่มมากขึ้นความเสี่ยงในการเกิดโรคต่าง ๆ รวมทั้งโรคมะเร็งจะเพิ่มสูงขึ้น [7]

เพศเพศหญิงมีแนวโน้มเป็นมะเร็งและเสียชีวิตจากมะเร็งมากกว่าเพศชายโดยเฉพาะผู้หญิงที่สูบบุหรี่, ดื่มแอลกอฮอล์ หรือทำงานในสถานที่ที่มีมลภาวะมาก ความเสี่ยงในการเป็นมะเร็งจะเพิ่มมากขึ้น นอกจากนี้การฉีดฮอร์โมนเอสโตรเจนทดแทนในหญิงวัยหมดประจำเดือน ก็ถือเป็นปัจจัยสำคัญที่เพิ่มโอกาสในเป็นมะเร็งด้วย [8,9]

เบาหวานโรคเบาหวานชนิดที่ 2 (Type 2 Diabetes) เกิดจากตับอ่อนสร้างอินซูลินได้ไม่เพียงพอต่อความต้องการของร่างกายหรือเกิดภาวะดื้อต่ออินซูลินซึ่งอาจทำให้ความเสี่ยงต่อการเป็นมะเร็งลำไส้ใหญ่เพิ่มขึ้น เนื่องจากผู้ป่วยจะมีระดับอินซูลินในเลือดสูงกว่าปกติ ซึ่งอินซูลินถือเป็น growth factor ตัวหนึ่ง ที่อาจกระตุ้นการเจริญของเซลล์และเร่งการพัฒนาจนเกิดเป็นเซลล์มะเร็งในที่สุด[10, 11, 12, 18]

การรับประทานอาหารที่มีไขมันสูงเช่น เนื้อแดง, เนื้อหมู, ผลิตภัณฑ์แปรรูปจากเนื้อสัตว์, ไข่แดง, นม และผลิตภัณฑ์แปรรูปจากนมไขมันประเภทนี้มีความเกี่ยวข้องกับโรคมะเร็งหลายชนิดเช่น มะเร็งลำไส้ใหญ่, มะเร็งทรวงอก, มะเร็งเยื่อบุโพรงมดลูก, มะเร็งรังไข่และมะเร็งต่อมลูกหมาก เพราะเมื่อนำอาหารประเภทนี้ไปปรุงในอุณหภูมิที่ร้อนจัดจะทำให้เกิดสารก่อมะเร็งขึ้น [13, 14]

ดื่มเครื่องดื่มที่มีแอลกอฮอล์มีหลักฐานมากมายที่แสดงให้เห็นได้อย่างชัดเจนว่าการดื่มเครื่องดื่มที่มีแอลกอฮอล์ เป็นปัจจัยสำคัญที่ทำให้เกิดโรคมะเร็งหลายชนิด โดยเฉพาะมะเร็งตับ, มะเร็งหลอดอาหาร และมะเร็งทางเดินอาหาร โดยจะทำให้เกิดโรคไขมันคั่งในตับ และโรคตับแข็งก่อนที่จะเป็นมะเร็งตับ เครื่องดื่มที่มีแอลกอฮอล์เป็นทูเมอร์โปรโมเตอร์ อัตราเสี่ยงการเกิดมะเร็งจะเพิ่มตามจำนวนครั้งของการดื่มและเปอร์เซ็นต์แอลกอฮอล์ในสุราที่ดื่ม [15, 16, 18]

มีประวัติคนในครอบครัวเป็นมะเร็งมะเร็งส่วนใหญ่ที่ถ่ายทอดกันทางพันธุกรรม ได้แก่ มะเร็งตับอ่อน, มะเร็งต่อมลูกหมาก, มะเร็งลำไส้ใหญ่, มะเร็งเต้านม และมะเร็งรังไข่ หากในครอบครัวมีประวัติการป่วยเป็นมะเร็งมีแนวโน้มที่บุคคลภายในครอบครัวจะเป็นมะเร็งด้วย[17]

การได้รับสารพิษและมลภาวะในสิ่งแวดล้อมการสัมผัสสารต่าง ๆ เช่น โลหะ ฝุ่น สารเคมี ยาฆ่าแมลงเป็นเวลานาน ๆ อาจก่อให้เกิดมะเร็งได้ ตัวอย่างของสารก่อมะเร็ง ได้แก่ นิโคติน แคลเมียม ยูเรเนียม เรดอน ไวนิลคลอไรด์ เบนซินดีน และเบนซิน โดยปกติสารก่อมะเร็งเหล่านี้จะต้องเข้าสู่ร่างกาย เช่น ผ่านการสูดดม หรือการสัมผัส เป็นเวลานาน กว่าจะทำให้เซลล์ปกติภายในร่างกายเกิดการกลายพันธุ์เซลล์มะเร็ง[18]

รังสี กระบวนการตรวจวินิจฉัยทางรังสีวิทยาอาจเพิ่มความเสี่ยงของการเกิดโรคมะเร็งได้ แม้จะได้รับในปริมาณเพียงเล็กน้อยแต่หากได้รับรังสีซ้ำกันหลาย ๆ ครั้ง ก็อาจก่อให้เกิดอันตรายได้[19]

การติดเชื้อไวรัสบางชนิด เช่น ไวรัสตับอักเสบบีและซี มีความสัมพันธ์ต่อการเกิดโรคตับแข็งและมะเร็งตับ ไวรัสเอชพีวี (Human Papiloma Virus)เป็นปัจจัยเสี่ยงของการเกิดโรคมะเร็งปากมดลูกในเพศหญิง [18]

การตากแดดจัดเป็นเวลานาน มีความเสี่ยงที่จะเป็นมะเร็งผิวหนังเพราะรังสีอัลตราไวโอเล็ตจะเข้าไปทำลายดีเอ็นเอ ซึ่งเป็นสาเหตุที่ทำให้เกิดมะเร็งผิวหนังอีกทั้งในแสงแดดยังมีสารกระตุ้นเซลล์มะเร็งดังนั้นหากยังปล่อยให้ผิวหนังสัมผัสกับแสงแดดจัดโดยตรงเป็นเวลานาน ความเสี่ยงในการเป็นมะเร็งผิวหนังก็จะเพิ่มมากขึ้น โดยกว่า 90% ของผู้ป่วยมะเร็งผิวหนังล้วนมีสาเหตุมาจากรังสีอัลตราไวโอเล็ต[20]

ในปัจจุบันยังไม่ทราบแน่ชัดว่าโรคมะเร็งมีสาเหตุมาจากอะไร จากการทบทวนวรรณกรรมและงานวิจัยพบว่า ปัจจัยที่มีความสัมพันธ์ต่อการเกิดโรคมะเร็งมีจำนวนมาก อาทิเช่นอายุ เพศ การสูบบุหรี่ การรับประทานผักผลไม้สด ภาวะระดับไขมันในเลือดสูง โรคเบาหวาน การดื่มแอลกอฮอล์ การขาดการออกกำลังกาย การได้รับควันบุหรี่มือสอง การติดเชื้อบางชนิด ประวัติการป่วยเป็นมะเร็งของคนในครอบครัว การได้รับสารพิษบางชนิด การตากแดดจัดเป็นประจำ ผู้ที่มีภูมิคุ้มกันบกพร่องอันเกิดจาก

ความผิดปกติทางพันธุกรรมหรือติดเชื้อไวรัสเอชไอวี ผู้ที่เป็นพาหะของเชื้อไวรัสตับอักเสบบี และผู้ที่ติดเชื้อพยาธิใบไม้ในตับ เป็นต้นเนื่องจากข้อจำกัดของข้อมูลที่นำมาศึกษาและจากศึกษาผลการวิจัยที่ผ่านมาของผู้วิจัยส่วนใหญ่ พบว่ามีตัวแปรที่สามารถคัดเลือกมาใช้ในการศึกษาได้เพียง 7 ตัวแปร ได้แก่ อายุ เพศ การสูบบุหรี่ การรับประทานผักผลไม้เพียงพอก ภาวะระดับไขมันในเลือดสูง โรคเบาหวาน และการป่วยเป็นโรคมะเร็งผู้วิจัยจึงใช้ตัวแปรทั้ง 7 ตัวแปรนี้ในการศึกษาความสัมพันธ์ระหว่างตัวแปรต่าง ๆ เหล่านี้กับการเป็นหรือไม่เป็นโรคมะเร็ง โดยลักษณะตัวแปรที่ใช้ในการสร้างสมการทำนายดังแสดงในตารางที่ 1

ตารางที่ 1.ลักษณะของตัวแปรที่ใช้ในงานวิจัย

ตัวแปร	รหัสตัวแปร
โรคมะเร็ง (CANCER)	0 = ไม่เป็น, 1 = เป็น
อายุ (AGE)	0 = น้อยกว่า 45 ปี, 1 = มากกว่าหรือเท่ากับ 45 ปี
เพศ (SEX)	0 = ชาย, 1 = หญิง
การสูบบุหรี่ (SMOKE)	0 = ไม่สูบบุหรี่, 1 = สูบบุหรี่
การรับประทานผักผลไม้เพียงพอก (FRUIT)	0 = ไม่เพียงพอก, 1 = เพียงพอก
ภาวะระดับไขมันในเลือดสูง (CHOLESTEROL)	0 = ไม่เป็น, 1 = เป็น
เบาหวาน (DIABETES)	0 = ไม่เป็น, 1 = เป็น

2.2 แหล่งข้อมูลประชากรและตัวอย่าง

การศึกษาค้นคว้าครั้งนี้ใช้ข้อมูลทุติยภูมิจากโครงการการสำรวจพฤติกรรมเสี่ยงโรคไม่ติดต่อและการบาดเจ็บ พ.ศ. 2558 ประชากรที่ใช้ในการวิจัยคือ ประชาชนที่มีอายุ 15-79 ปี จำนวนทั้งสิ้น 9,344 ราย ในจำนวนนี้มีผู้ป่วยโรคมะเร็ง 84 ราย และผู้ที่ไม่เป็นโรคมะเร็ง 9,260 ราย ซึ่งเห็นได้ว่าข้อมูลมีการกระจายที่เบ้มากมีความไม่สมดุลของจำนวนผู้ป่วยที่เป็นและไม่เป็นมะเร็ง เพื่อปรับปรุงมาตรฐานความถูกต้อง (accuracy) ในการจำแนกในกรณีที่มีข้อมูลในแต่ละกลุ่ม (class) แตกต่างกันอย่างมาก (class-imbalanced data) วิธีหนึ่งที่นิยมใช้โดยไม่ต้องเปลี่ยนแปลงตัวแบบการจำแนกคือ วิธีการสุ่มลด (Under Sampling) เพื่อสุ่มเลือกข้อมูลของกลุ่มที่มีจำนวนมากให้มีจำนวนใกล้เคียงกับจำนวนข้อมูลของกลุ่มที่มีจำนวนน้อย [21] ทำให้ได้จำนวนข้อมูลกลุ่มหลัก (ไม่เป็นโรคมะเร็ง) และข้อมูลกลุ่มรอง (เป็นโรคมะเร็ง) ใกล้เคียงกัน ในการศึกษาค้นคว้าครั้งนี้ผู้วิจัยได้ทำการสุ่มลดข้อมูลกลุ่มหลักเพื่อสร้างชุดข้อมูล 3 ชุด คือชุดที่ 1 (ผู้ป่วยมะเร็ง : ผู้ที่ไม่เป็นมะเร็ง = 1 : 1), ชุดที่ 2 (ผู้ป่วยมะเร็ง : ผู้ที่ไม่เป็นมะเร็ง = 1 : 2) และชุดที่ 3 (ผู้ป่วยมะเร็ง : ผู้ที่ไม่เป็นมะเร็ง = 1 : 3) [22, 23, 24] เพื่อเปรียบเทียบประสิทธิภาพของสมการทำนายความน่าจะเป็นในการเกิดโรคมะเร็งที่สร้างจากข้อมูลชุดต่าง ๆ

2.3 การพัฒนาตัวแบบทำนายโอกาสการเป็นหรือไม่เป็นโรคมะเร็ง

ผู้วิจัยได้ทำการสร้างตัวแบบและเปรียบเทียบสมการเพื่อทำนายความน่าจะเป็นในการเป็นหรือไม่เป็นโรคมะเร็ง โดยมีการแบ่งข้อมูลทั้ง 3 ชุดออกเป็นข้อมูลเรียนรู้ (Training Data) 60% และเป็นข้อมูลทดสอบ (Validation Data) 40% [25]การคัดเลือกสมการทำนายใช้เกณฑ์ค่าสถิติจากตารางการจำแนก ได้แก่ ร้อยละความถูกต้อง , ร้อยละความไว, ร้อยละความเฉพาะ, ร้อยละบวกเท็จ, ร้อยละลบเท็จและร้อยละ

ผลการจำแนกผิด [3, 23, 25, 26] และกราฟเส้นโค้ง ROC ของแต่ละสมการ สมการทำนายที่มีประสิทธิภาพสูงคือสมการที่มีอัตราการทำจำแนกผิดน้อย มีค่าร้อยละความถูกต้องและพื้นที่ใต้โค้ง ROC มาก

2.4 การวิเคราะห์ข้อมูล(Data Analysis)

2.4.1 สถิติเชิงพรรณนา(Descriptive statistics)เป็นการวิเคราะห์เพื่ออธิบายข้อมูลเบื้องต้นเกี่ยวกับตัวแปรต่าง ๆ ที่ใช้ในงานวิจัยและตัวอย่าง โดยใช้ความถี่ ร้อยละ และค่าเฉลี่ย

2.4.2 การวิเคราะห์สองตัวแปร (Bivariate Analysis)

คำนวณความแตกต่างระหว่างตัวอย่างที่เป็นมะเร็งและไม่เป็นมะเร็ง ว่ามีลักษณะของเพศ อายุ โรคเบาหวาน ภาวะระดับไขมันในเลือดสูง การรับประทานผักผลไม้เพียงพอก และสูบบุหรี่แตกต่างกันอย่างไร โดยใช้สถิติทดสอบไคกำลังสอง (χ^2 - test) [13, 26, 28]

2.4.3 การวิเคราะห์การถดถอยลอจิสติก (Logistic regression)

เป็นการวิเคราะห์เพื่อศึกษาความสัมพันธ์และอิทธิพลของตัวแปรอิสระที่มีต่อตัวแปรตาม (เป็นและไม่เป็นโรคมะเร็ง) โดยคัดเลือกตัวแปรอิสระเข้าสู่ตัวแบบการถดถอยลอจิสติกด้วยวิธีการถดถอยแบบขั้นตอน (Stepwise) กำหนดระดับนัยสำคัญที่จะนำตัวแปรอิสระเข้าสู่ตัวแบบและนำออกจากตัวแบบเท่ากับ 0.05 และ 0.1 ตามลำดับ ตรวจสอบความเหมาะสมของตัวแบบลอจิสติกที่สร้างขึ้นด้วยสถิติทดสอบฮอสเมอร์และเลเมโชว์ (Hosmer and Lemeshow)[16]และสถิติจากตารางการจำแนก (Classification table)สำหรับการวินิจฉัยว่าสมการถดถอยลอจิสติกสร้างมาจากตัวแบบที่เหมาะสมหรือไม่ จะพิจารณาถึงอิทธิพลของค่าสังเกต โดยเกณฑ์ที่ใช้ในการตรวจสอบ ได้แก่ส่วนเหลือดีวีเนียนซ์ (Deviance residual) และสมาชิกทแยงมุมของเมทริกซ์หมวก(Hat matrix diagonal)[3,13,26,28,29]

3. ผลการวิเคราะห์ข้อมูล

ข้อมูลเบื้องต้นจากตัวอย่างพบว่า เพศหญิงมีร้อยละ 64.29 ซึ่งมากกว่าเพศชายที่มีร้อยละ 35.71 ตามลำดับ เพศหญิงเป็นมะเร็งร้อยละ 30.56 เพศชายเป็นมะเร็งร้อยละ 15.00 ประชาชนที่มีอายุมากกว่าหรือเท่ากับ 45 ปี ร้อยละ 83.04ในจำนวนนี้มีผู้เป็นโรคมะเร็งร้อยละ 28.67 ส่วนผู้ที่มีอายุน้อยกว่า 45 ปีที่เป็นมะเร็งร้อยละ 7.02 มีประชาชนเป็นโรคเบาหวานคิดเป็นร้อยละ 16.07ในจำนวนนี้มีผู้ที่เป็นมะเร็งร้อยละ 35.19 ส่วนผู้ที่ไม่เป็นโรคเบาหวานเป็นมะเร็งร้อยละ 23.05 จากตัวอย่างผู้ที่มีภาวะระดับไขมันในเลือดสูงมีร้อยละ 28.57 ในจำนวนนี้เป็นมะเร็งร้อยละ 31.25 ผู้ที่ไม่มีภาวะระดับไขมันในเลือดสูงมีร้อยละ 71.43 ในจำนวนนี้เป็นมะเร็งร้อยละ 22.50ผู้ที่รับประทานผักผลไม้เพียงพอมีร้อยละ 70.24มีผู้เป็นมะเร็งร้อยละ 26.27 ส่วนผู้ที่รับประทานผักผลไม้ไม่เพียงพอเป็นมะเร็งมีร้อยละ 22.00 และมีผู้สูบบุหรี่ร้อยละ 23.51ในจำนวนนี้เป็นมะเร็งร้อยละ 34.18 ส่วนผู้ที่ไม่สูบบุหรี่จะเป็นมะเร็งร้อยละ 22.18และจากการทดสอบความสัมพันธ์ระหว่างตัวแปรอิสระกับตัวแปรตาม (เป็นและไม่เป็นโรคมะเร็ง)โดยใช้สถิติทดสอบไคกำลังสองพบว่า มีตัวแปรอิสระเพียง 3 ตัวแปรเท่านั้นคือ เพศ อายุ และการสูบบุหรี่ ที่มีความสัมพันธ์ต่อการเกิดโรคมะเร็งอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 โดยค่าไคกำลังสองของตัวแปรเพศ อายุ และการสูบบุหรี่เท่ากับ 9.9556, 11.8388 และ 4.6393 ตามลำดับ และมีค่าพีเท่ากับ 0.0016, 0.0006 และ 0.0312 ตามลำดับ ดังรายละเอียดในตารางที่ 2

ตารางที่ 2. การทดสอบความแตกต่างระหว่างตัวอย่างที่เป็นและไม่เป็นมะเร็งกับเพศ อายุ โรคเบาหวาน ภาวะระดับไขมันในเลือดสูง การรับประทานผักผลไม้อย่างเพียงพอและการสูบบุหรี่ (n=336)

ตัวอย่าง	ร้อยละ	ร้อยละ		ค่าไคกำลังสอง	องศาเสรี	ค่าพี
		เป็นมะเร็ง	ไม่เป็นมะเร็ง			
เพศ				9.9556	1	0.0016**
ชาย	35.71	15.00	85.00			
หญิง	64.29	30.56	69.44			
อายุ				11.8388	1	0.0006**
< 45 ปี	16.96	7.02	92.98			
≥ 45 ปี	83.04	28.67	71.33			
เบาหวาน				3.5598	1	0.0592
เป็น	16.07	35.19	64.81			
ไม่เป็น	83.93	23.05	76.95			
ภาวะระดับไขมันในเลือดสูง				2.8000	1	0.0943
เป็น	28.57	31.25	68.75			
ไม่เป็น	71.43	22.50	77.50			
การรับประทานผักผลไม้อย่างเพียงพอ				0.6834	1	0.4184
เพียงพอ	70.24	26.27	73.73			
ไม่เพียงพอ	29.76	22.00	78.00			
การสูบบุหรี่				4.6393	1	0.0312**
สูบบุหรี่	23.51	34.18	65.82			
ไม่สูบบุหรี่	76.49	22.18	77.82			

หมายเหตุ ** มีนัยสำคัญที่ระดับ 0.05

3.1 ผลการคัดเลือกสมการทำนายความน่าจะเป็นในการเป็นมะเร็ง

จากตารางที่ 2 พบว่า ตัวแปรอิสระที่มีความสัมพันธ์ต่อการเกิดโรคมะเร็งอย่างมีนัยสำคัญทางสถิติ มีเพียง 3 ตัวแปรเท่านั้น คือ เพศ อายุ และการสูบบุหรี่ ผู้วิจัยจึงนำตัวแปรเหล่านี้ไปวิเคราะห์การถดถอยโลจิสติก ผลการวิเคราะห์ดังแสดงในตารางที่ 4 และใช้เกณฑ์ค่าสถิติจากตารางการจำแนก เพื่อเปรียบเทียบประสิทธิภาพของสมการทำนายความน่าจะเป็นในการเกิดโรคมะเร็งที่สร้างขึ้นจากข้อมูลชุดต่าง ๆ

จากตารางที่ 3 จะได้ว่า ข้อมูลชุดที่ 1 มีค่าร้อยละความถูกต้องเท่ากับ 58.90 ร้อยละความไวเท่ากับ 97.60 ร้อยละความเฉพาะเท่ากับ 20.20 ร้อยละบวกเท็จเท่ากับ 45.00 ร้อยละลบเท็จเท่ากับ 10.50 และร้อยละการจำแนกผิดเท่ากับ 41.10 และพื้นที่ใต้โค้ง ROC เท่ากับ 0.6420 ข้อมูลชุดที่ 2 มีค่าร้อยละความถูกต้องเท่ากับ 69.80 ร้อยละความไวเท่ากับ 13.10 ร้อยละความเฉพาะเท่ากับ 98.20 ร้อยละบวกเท็จเท่ากับ 21.40 ร้อยละลบเท็จเท่ากับ 30.70 ร้อยละการจำแนกผิดเท่ากับ 30.20 และพื้นที่ใต้โค้ง ROC เท่ากับ 0.6805 ข้อมูลชุดที่ 3 มีค่าร้อยละความถูกต้องเท่ากับ 76.80 ร้อยละความไวเท่ากับ 13.10 ร้อยละความเฉพาะเท่ากับ 98.00 ร้อยละบวกเท็จเท่ากับ 31.30 ร้อยละลบเท็จเท่ากับ 22.80 ร้อยละการจำแนกผิดเท่ากับ 23.20 และพื้นที่ใต้โค้ง ROC เท่ากับ 0.6964

ตารางที่ 3. ร้อยละการทำนายของตัวแบบลอจิสติกกับการเป็นหรือไม่เป็นมะเร็งของตัวอย่าง

ค่าสังเกตการ เป็นมะเร็ง	ข้อมูลชุดที่ 1			ข้อมูลชุดที่ 2			ข้อมูลชุดที่ 3		
	เป็นมะเร็ง : ไม่เป็นมะเร็ง = 1:1			เป็นมะเร็ง : ไม่เป็นมะเร็ง 1:2			เป็นมะเร็ง : ไม่เป็นมะเร็ง 1:3		
	สมการทำนายการเป็นมะเร็ง			สมการทำนายการเป็นมะเร็ง			สมการทำนายการเป็นมะเร็ง		
	ไม่เป็น	เป็น	รวม	ไม่เป็น	เป็น	รวม	ไม่เป็น	เป็น	รวม
ไม่เป็น	17	67	84	165	3	168	247	5	252
เป็น	2	82	84	73	11	84	73	11	84
รวม	19	149	168	238	14	252	320	16	336

ตารางที่ 4 แสดงค่าประมาณพารามิเตอร์ของตัวแบบถดถอยลอจิสติกที่ได้จากข้อมูลชุดที่ 3 (ผู้ป่วยมะเร็ง : ผู้ป่วยที่ไม่เป็นมะเร็ง = 1:3) ซึ่งเป็นตัวแบบการถดถอยลอจิสติกที่มีประสิทธิภาพสูงสุดคือ มีร้อยละความถูกต้องมากที่สุด มีอัตราการทำนายที่น้อยที่สุด และมีพื้นที่ใต้โค้ง ROC มากที่สุด โดยตัวแบบการถดถอยลอจิสติกที่สร้างขึ้นสามารถจำแนกตัวอย่างที่เป็นมะเร็งและไม่เป็นมะเร็งได้ถูกต้องร้อยละ 76.80 จำแนกตัวอย่างที่เป็นมะเร็งได้ถูกต้องร้อยละ 13.10 จำแนกตัวอย่างที่ไม่เป็นมะเร็งได้ถูกต้องร้อยละ 98.00 มีร้อยละบวกเท็จ (ความน่าจะเป็นของความผิดพลาดชนิดที่ 1) เท่ากับ 31.30 หมายความว่าสมการสามารถจำแนกตัวอย่างที่ไม่เป็นมะเร็งว่าเป็นมะเร็งร้อยละ 31.30 ร้อยละลบเท็จ (ความน่าจะเป็นของความผิดพลาดชนิดที่ 2) เท่ากับ 22.80 หมายความว่าสมการสามารถจำแนกตัวอย่างที่เป็นมะเร็งว่าไม่เป็นมะเร็งร้อยละ 22.80

ตารางที่ 4. ค่าประมาณพารามิเตอร์ของตัวแบบการถดถอยลอจิสติก

ตัวแปร	ค่าประมาณพารามิเตอร์	ความคลาดเคลื่อนมาตรฐาน	สถิติวอลด์	ค่าพี	อัตราส่วนออดส์	ช่วงความเชื่อมั่น 95 % ของอัตราส่วนออดส์
ส่วนตัด	-4.3277	0.6664	42.1782	<0.0001		
เพศ (SEX)	1.9939	0.4591	19.3669	<0.0001**	7.3440	3.0220 - 17.8490
อายุ (AGE)	1.5339	0.5411	8.0344	0.0046**	4.6360	1.6050 - 13.3900
การสูบบุหรี่ (SMOKE)	1.8171	0.4480	16.4552	<0.0001**	6.1540	2.5580 - 14.8090

หมายเหตุ ** มีนัยสำคัญที่ระดับ 0.05

สมการถดถอยลอจิสติก คือ

$$\hat{P}[\text{CANCER} = 1|\mathbf{x}] = \frac{e^{\hat{g}(\mathbf{x})}}{1+e^{\hat{g}(\mathbf{x})}} \quad (1)$$

$$\text{โดยที่ } \hat{g}(\mathbf{x}) = -4.3277 + 1.9939\text{SEX} + 1.5339\text{AGE} + 1.8171\text{SMOKE} \quad (2)$$

จากการทดสอบพารามิเตอร์ของตัวแบบทุกตัวพบว่าสถิติทดสอบอัตราส่วนภาวะน่าจะเป็น มีค่าไคกำลังสอง = 45.1037 และค่า $p < 0.0001$ แสดงว่าสัมประสิทธิ์ของลอจิสติกในสมการ (2) แตกต่างจากศูนย์อย่างมีนัยสำคัญ แสดงว่าตัวแปรอิสระทั้ง 3 ตัว คือ เพศ อายุ และการสูบบุหรี่ มีอิทธิพลต่อการเป็นมะเร็งหรือไม่เป็นมะเร็งของตัวอย่าง

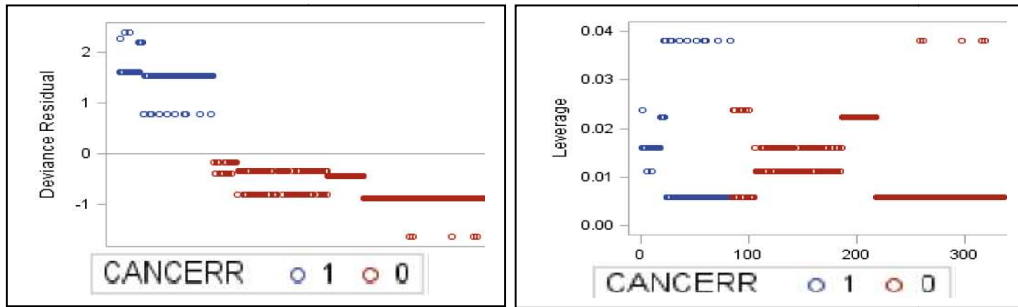
จากการทดสอบพารามิเตอร์แต่ละตัวของตัวแบบโดยใช้ค่าสถิติวอลด์ (Wald statistics) พบว่า ตัวแปรอิสระที่มีผลต่อตัวแปรมะเร็งอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 มี 3 ตัว ได้แก่ เพศ (SEX) อายุ (AGE) และการสูบบุหรี่ (SMOKE) โดยค่าสถิติวอลด์ของตัวแปรเพศ อายุ และการสูบบุหรี่ เท่ากับ 19.3669, 8.0344 และ 16.4552 ตามลำดับ และมีค่า p เท่ากับ < 0.0001 , 0.0046 และ < 0.0001 ตามลำดับ แสดงว่าสัมประสิทธิ์การถดถอยลอจิสติกของตัวแปรเพศ อายุ และการสูบบุหรี่ แต่ละตัวแตกต่างจากศูนย์อย่างมีนัยสำคัญ

จากตารางที่ 4 พบว่า ตัวแปรเพศมีค่าอัตราส่วนออดส์เท่ากับ 7.3440 หมายความว่า ในประชากรที่ทำการศึกษเพศหญิงมีความเสี่ยงที่จะเป็นโรคมะเร็งมากกว่าเพศชาย 7.3440 เท่า หรือเพศหญิงมีความเสี่ยงที่จะเป็นโรคมะเร็งมากกว่าเพศชายเท่ากับ 3.0220 เท่าถึง 17.8490 เท่าด้วยความเชื่อมั่น 95% ตัวแปรอายุมีค่าอัตราส่วนออดส์เท่ากับ 4.6360 หมายความว่า ในประชากรที่ทำการศึกษาผู้ที่มีอายุมากกว่าหรือเท่ากับ 45 ปีมีความเสี่ยงที่จะเป็นโรคมะเร็งมากกว่าผู้ที่มีอายุน้อยกว่า 45 ปี 4.6360 เท่า หรือผู้ที่มีอายุมากกว่าหรือเท่ากับ 45 ปีมีความเสี่ยงที่จะเป็นโรคมะเร็งมากกว่าผู้ที่มีอายุน้อยกว่า 45 ปีเท่ากับ 1.6050 เท่าถึง 13.3900 เท่าด้วยความเชื่อมั่น 95% และตัวแปรการสูบบุหรี่มีค่าอัตราส่วนออดส์เท่ากับ 6.1540 หมายความว่า ประชากรที่ทำการศึกษาคคนที่สูบบุหรี่มีความเสี่ยงที่จะเป็นโรคมะเร็งมากกว่าคนที่ไม่สูบบุหรี่ 6.1540 เท่า หรือคนที่สูบบุหรี่มีความเสี่ยงที่จะเป็นโรคมะเร็งมากกว่าคนที่ไม่สูบบุหรี่เท่ากับ 2.5580 เท่าถึง 14.8090 เท่าด้วยความเชื่อมั่น 95%

3.2 ผลการตรวจสอบความเหมาะสมของตัวแบบลอจิสติก

ในการตรวจสอบว่าตัวแบบที่ใช้ในการศึกษานี้มีความเหมาะสมกับข้อมูลเชิงประจักษ์หรือไม่ โดยใช้การทดสอบของฮอสเมอร์และเลมโชว์ (Hosmer and Lemeshow Test) ซึ่งพบว่า การทดสอบของฮอสเมอร์และเลมโชว์มีค่าไคกำลังสองเท่ากับ 0.2464 และค่า p เท่ากับ 0.8841 ที่ระดับนัยสำคัญ 0.05 จึงสรุปได้ว่าตัวแบบการถดถอยลอจิสติกที่สร้างขึ้นมีความเหมาะสมกับข้อมูลเชิงประจักษ์

การพิจารณาค่านอกเกณฑ์ของค่าสังเกตที่มีผลกระทบในการวิเคราะห์การถดถอยลอจิสติก โดยพิจารณาจากค่าสัมบูรณ์ของส่วนเหลือดีวีเนียนซ์มีค่าไม่เกิน 3 แสดงว่าไม่มีค่าสังเกตใดที่มีแนวโน้มจะเป็นค่าที่มีอิทธิพลต่อสมการถดถอยลอจิสติก ดังแสดงในภาพที่ 1. จากการตรวจสอบค่าเลฟเวอร์ริจ (leverage values) พบว่า ไม่มีค่าสังเกตใดที่มีค่าเลฟเวอร์ริจสูง แสดงว่าไม่มีค่าสังเกตที่มีแนวโน้มที่จะเป็นค่าที่มีอิทธิพลต่อสมการถดถอยลอจิสติก ดังแสดงในภาพที่ 1.ข



ภาพที่ 1. ส่วนเหลือดีเวียนซ์ภาพที่ 1. ขสมากิตกแยงมมของเมทริกซ์หวมก

4. สรุปผลการวิจัย

การศึกษาเพื่อหาปัจจัยเสี่ยงที่มีความสัมพันธ์ต่อการเกิดโรคมะเร็ง โดยการสร้างสมการถดถอยลอจิสติกเนื่องจากข้อมูลที่น่าสนใจในงานวิจัยมีการกระจายที่เบ้มาก มีความไม่สมดุลของจำนวนผู้ป่วยที่เป็นและไม่เป็นมะเร็ง เพื่อการปรับปรุงมาตรวัดความถูกต้อง และเพื่อป้องกันความเอนเอียงของผลการวิเคราะห์จึงทำการปรับข้อมูลให้สมดุลด้วยวิธีการสุ่มลด เพื่อให้จำนวนข้อมูลในกลุ่มที่ไม่เป็นโรคมะเร็งและในกลุ่มที่มะเร็งใกล้เคียงกัน โดยทำการสุ่มลดข้อมูลกลุ่มหลัก (ไม่เป็นมะเร็ง) เพื่อสร้างชุดข้อมูล 3 ชุด เพื่อเปรียบเทียบประสิทธิภาพของตัวแบบที่สร้างจากข้อมูลชุดต่าง ๆ โดยใช้เกณฑ์ค่าสถิติจากตารางการจำแนก จากการวิเคราะห์ข้อมูลพบว่าตัวแบบที่มีประสิทธิภาพสูงสุดคือตัวแบบที่สร้างมาจากข้อมูลชุดที่ 3 (ผู้ป่วยมะเร็ง : ผู้ไม่เป็นมะเร็ง = 1 : 3)

จากการวิเคราะห์ข้อมูล พบว่า ตัวแปรอิสระที่มีความสัมพันธ์ต่อการเกิดโรคมะเร็งอย่างมีนัยสำคัญทางสถิติมีเพียง 3 ตัวแปรเท่านั้น คือ เพศ อายุ และการสูบบุหรี่ และเมื่อนำตัวแปรเหล่านี้ไปวิเคราะห์การถดถอยลอจิสติก จะได้สมการถดถอยลอจิสติก ดังนี้

$$\hat{P}[\text{CANCER} = 1 | \mathbf{x}] = \frac{e^{\hat{g}(\mathbf{x})}}{1 + e^{\hat{g}(\mathbf{x})}} \quad (1)$$

$$\text{โดยที่ } \hat{g}(\mathbf{x}) = -4.3277 + 1.9939\text{SEX} + 1.5339\text{AGE} + 1.817\text{ISMOKE} \quad (2)$$

จากการวิเคราะห์ข้อมูลพบว่า สัมประสิทธิ์ของลอจิสติกในสมการ (2) ทุกตัว แตกต่างจากศูนย์อย่างมีนัยสำคัญที่ระดับนัยสำคัญ 0.05 ด้วยค่าไคกำลังสอง = 45.1037 และค่าพี < 0.0001 นั่นคือตัวแปรอิสระทั้ง 3 ตัว คือ เพศ อายุ และการสูบบุหรี่ มีอิทธิพลต่อการเป็นมะเร็งหรือไม่เป็นมะเร็งของตัวอย่าง และสัมประสิทธิ์การถดถอยลอจิสติกของตัวแปรเพศ อายุ และการสูบบุหรี่ แต่ละตัวแตกต่างจากศูนย์อย่างมีนัยสำคัญที่ระดับ 0.05 โดยค่าสถิติวอลด์ของตัวแปรเพศ อายุ และการสูบบุหรี่ เท่ากับ 19.3669, 8.0344 และ 16.4552 ตามลำดับ และมีค่าพีเท่ากับ < 0.0001, 0.0046 และ < 0.0001 ตามลำดับ

ผลการวิจัยสรุปได้ว่าตัวแบบที่สร้างขึ้นมีความเหมาะสมกับข้อมูลเชิงประจักษ์ ด้วยค่าไคกำลังสองเท่ากับ 0.2464 และค่าพีเท่ากับ 0.8841 ที่ระดับนัยสำคัญ 0.05 และสมการที่สร้างขึ้นมีความถูกต้องร้อยละ 76.80, ความไวร้อยละ 13.10, ความเฉพาะร้อยละ 98.00, อัตราบวกเท็จ (ความน่าจะเป็นของความ

ผลิตผลชนิดที่ 1) ร้อยละ 31.30 และอัตราลบเท็จ (ความน่าจะเป็นของความผิดพลาดชนิดที่ 2) ร้อยละ 22.80 เมื่อกำหนดความน่าจะเป็นในการแบ่งกลุ่มเท่ากับ 0.5 และเมื่อพิจารณาอัตราส่วนอัตราส่วนของตัวแปรอิสระแต่ละตัวพบว่า ตัวแปรอิสระที่ทำนายความน่าจะเป็นของการเป็นมะเร็งในรูปอัตราส่วนอัตราส่วนได้มากที่สุดคือ เพศ รองลงมาได้แก่ การสูบบุหรี่ อายุ ตามลำดับ โดยเพศหญิงมีความเสี่ยงที่จะเป็นมะเร็งมากกว่าเพศชาย 7.3440 เท่า คนสูบบุหรี่มีความเสี่ยงที่จะเป็นมะเร็งมากกว่าคนที่ไม่สูบบุหรี่ 6.1540 เท่า และคนที่มียาอายุมากกว่าหรือเท่ากับ 45 ปีมีความเสี่ยงที่จะเป็นมะเร็งมากกว่าคนที่มียาอายุน้อยกว่า 45 ปี 4.6360

ผลการวินิจฉัยสมการถดถอยลอจิสติกว่าสร้างมาจากตัวแบบที่เหมาะสมหรือไม่นั้น จะพิจารณาถึงอิทธิพลของค่าสังเกตว่ามีผลกระทบอย่างไรในการวิเคราะห์การถดถอยลอจิสติก จากการตรวจสอบพบว่าค่าสัมบูรณ์ของส่วนเหลือดีเวียนซ์มีค่าไม่เกิน ± 3 และไม่มีค่าสังเกตใดมีค่าเลฟเวอร์ริงสูง แสดงว่าไม่มีค่าสังเกตใดที่มีแนวโน้มจะเป็นค่าที่มีอิทธิพลต่อสมการถดถอยลอจิสติก

5. ข้อเสนอแนะ

1. การปรับเปลี่ยนพฤติกรรมเสี่ยงที่คาดว่าจะมีความสัมพันธ์กับการเกิดโรคมะเร็ง เช่น การสูบบุหรี่ การดื่มเครื่องดื่มที่มีแอลกอฮอล์นั้น เป็นเรื่องที่รัฐบาลจะต้องทำให้เกิดเป็นรูปธรรม มีช่องทางการเผยแพร่ข้อมูล โฆษ และความร้ายแรงของโรคมะเร็งที่ถูกต้องผ่านสื่อต่าง ๆ ให้เกิดความเข้าใจแก่ประชาชนในระดับอย่างทั่วถึง เพื่อป้องกันและลดปัญหาโรคมะเร็งในประเทศไทย

2. การศึกษาการเป็นหรือไม่เป็นโรคมะเร็งนั้น ถือเป็นปัญหาสุขภาพที่สำคัญของประชาชนในประเทศไทย และเป็นโรคที่มีอัตราการตายสูงสุดในแต่ละปี จึงเห็นว่รัฐบาลควรให้การสนับสนุนการศึกษาในด้านนี้ โดยเฉพาะเรื่องงบประมาณในการทำวิจัย เพราะการวิจัยประเภทนี้มีค่าใช้จ่ายในการทำทดลองในห้องปฏิบัติการ การเก็บรวบรวมข้อมูลและการวิเคราะห์ข้อมูลค่อนข้างสูง

3. จากผลการวิจัยพบว่าคนสูบบุหรี่มีความเสี่ยงที่จะป่วยเป็นโรคมะเร็งมากกว่าคนที่ไม่สูบบุหรี่ นอกจากนี้ยังมีงานวิจัยทางการแพทย์ที่แสดงให้เห็นได้ว่า ไม่ใช่เฉพาะผู้สูบบุหรี่เท่านั้นที่จะได้รับผลกระทบด้านสุขภาพ แต่คนรอบข้างที่รับเอาควันบุหรี่เข้าไป หรือที่เรียกว่า “ควันบุหรี่มือสอง” ก็จะได้รับผลกระทบด้านสุขภาพด้วย ดังนั้นหน่วยงานที่รับผิดชอบ จึงควรจัดให้มีการเผยแพร่ข้อมูลเกี่ยวกับอันตรายจากควันบุหรี่มือสอง และปัญหาสุขภาพที่อาจเกิดขึ้นจากการได้รับควันบุหรี่มือสองนอกจากนี้ควรส่งเสริมให้มีการตรวจสุขภาพประจำปีของผู้ได้รับควันบุหรี่มือสอง ในรูปแบบที่สะดวก ประหยัด และเสียเวลาน้อยเพื่อเป็นการคัดกรองและป้องกันการเจ็บป่วยจากการได้รับควันบุหรี่มือสอง

กิตติกรรมประกาศ

ขอขอบคุณ ดร.กมลทิพย์ วิจิตรสุนทรกุล ผู้สนับสนุนข้อมูลในการทำวิจัย ผู้วิจัยขอขอบพระคุณเป็นอย่างยิ่ง

เอกสารอ้างอิง (References)

- [1] สุพรรณี พรหมเทศ และคณะ. 2553. ปัจจัยเสี่ยงต่อการเกิดโรคมะเร็งในคนไทยภาคตะวันออกเฉียงเหนือ. รายงานวิจัย. [Supanee Promtech and et.al. 2010. Risk Factors for Cancer in Northeastern Region in Thailand. Research Report. (in Thai)]
- [2] คณะกรรมการจัดทำแผนการป้องกันและควบคุมโรคมะเร็งแห่งชาติ. 2556. แผนการป้องกันและควบคุมโรคมะเร็งแห่งชาติ พ.ศ. 2558 - 2560. พิมพ์ครั้งที่ 1, โรงพิมพ์ชุมนุมสหกรณ์การเกษตรแห่งประเทศไทย จำกัด, กรุงเทพมหานคร. [The National Cancer Prevention and Control Commission. 2013. National Cancer Control Programmes. 1sted, Thw Agricultural Co-operative Federatiob of Thailand., LTD. Bangkok (in Thai)]
- [3] จิราวัลย์ จิตรถเวช. 2558. การวิเคราะห์การถดถอย. พิมพ์ครั้งที่ 1, โรงพิมพ์องค์การสงเคราะห์ทหารผ่านศึกในพระบรมราชูปถัมภ์, กรุงเทพฯ. [JirawanJitthavech. 2015. Regression Analysis. 1sted, The War Veterans Organization of Thailand Under Royal Patronage of His Majesty the King Publisher. Bangkok. (in Thai)]
- [4] Sarah, P. 2005. Gender in lung cancer and smoking research. Available from: <http://www.who.int/gender/documents/LungCancerlast2.pdf>. cited 2017 15 April.
- [5] Key, T.J.Schatzkin, A. Willett, W.C. Allen, N.E. Spencer, E.A. and Travis, R.C.2004. Diet, nutrition and the prevention of cancer. Public Health Nutrition, 7(1A), 187–200.
- [6] Oguntibeju, O.O. Truter, E.J. and Esterhuyse, A.J.2013. The Role of Fruit and Vegetable Consumption in Human Health and Disease Prevention. Available from: http://cdn.intechopen.com/pdfs/42095/InTechThe_role_of_fruit_and_vegetable_consumption_in_human_health_and_disease_prevention.pdf. cited 2017 5 March.
- [7] White, M.C.Holman, D.M. Boehm, J.E. Peipins, L.A. Grossman, M. and Henley, S.J. 2014. Age and Cancer Risk: A Potentially Modifiable Relationship: *American Journal of Preventive Medicine*, 46(3), 7–15.
- [8] Dorak, M.T. Karpuzoglu, E. 2012. Gender Differences in Cancer Susceptibility: An Inadequately Addressed Issue. *Frontiers in Genetics*, 268(3).
- [9] McCann, J. 2000. Gender Differences in Cancer That Don't Make Sense Or Do They?: *Journal of the National Cancer Institute*, 92(19), 1560-1562.
- [10] Cannata,D. Fierz, Y.Vijayakumar, A.andLeRoith, D. 2010. Type 2 Diabetes and Cancer: What Is the Connection?.*The Mount Sinai JournalOf Medicine*, 77(2), 197–213.
- [11] Chen, Y. Wu, F. Saito, E. and et al. 2017. Association between type 2 diabetes and risk of cancer mortality : a pooled analysis of over 771,000 individuals in the Asia Cohort Consortium. *Diabetologia*, 60(6), 1022-1032.

- [12] Silverman, D. T.Schiffman, M. and et al. 1999. Diabetes mellitus, other medical conditions and familial history of cancer as risk factors for pancreatic cancer. *British Journal of Cancer*,80(11), 1830-1837
- [13] ชัยยุทธ เจตียนูวัตร. 2544. ปัจจัยที่มีความสัมพันธ์กับการเลิกบุหรี่ของเจ้าหน้าที่สาธารณสุขจังหวัดกาญจนบุรี. วิทยานิพนธ์ปริญญาโทบริหารธุรกิจ, สาขาชีวสถิติ, มหาวิทยาลัยมหิดล. [ChaiyuthJetiyanuwatr. 2001. Factors associated with smoking cessation of public health officials in Kanchanaburi province. Thesis, Biostatistics Program, Mahidol University. (in Thai)]
- [14] La Vecchia, C. 1992. Cancers associated with high-fat diets.*Journal of National Cancer Institute Monograph*,(12), 79-85.
- [15] Connor, J. 2016. Alcohol consumption as a cause of cancer. *Addiction*, 112(2), 222-228.
- [16] Seitz, H.K, and Becker, P. 2007. Alcohol Metabolism and cancer risk. *Alcohol Research & Health*, 30(1), 38-47.
- [17] Jacobs, E.J. Chanock, S.J. and et al.2010. Family history of cancer and risk of pancreatic cancer: a pooled analysis from the Pancreatic Cancer Cohort Consortium (PanScan). *International Journal of Cancer*, 127(6), 1421-1428.
- [18] Linet, M.S.Slovis, T.L. Miller, D.L. Kleinerman, R. Lee, C. Rajaraman, P. and Berrington de Gonzalez A. 2012. Cancer Risks Associated With External Radiation From Diagnostic Imaging Procedures. *Cancer Journal for clinicians*, 62(2), 75-100.
- [19] Mohammed, A. Hannan, M.P.Magid H.A. and Farouk H.A. 1984. Study of Ultraviolet Radiation and Genotoxic Effects of Natural Sunlight in Relation to Skin Cancer in Saudi Arabia. *Cancer Research*, 44, 2192-2197.
- [20] ปภากร อินทร์โหล. 2557. การวิเคราะห์การถดถอยโลจิสติกสำหรับการดื้อยาต้านจุลชีพของเชื้อ *Salmonella enterica* ในสุกร. วิทยานิพนธ์ปริญญาโทบริหารธุรกิจ, สาขาสถิติ, มหาวิทยาลัยเกษตรศาสตร์.[PaphakornIntholo. 2014. Logistic Regression Analysis for Anyimicrobial Resistance of *Salmonella enterica* in swine. Thesis, Statistics Program, Kasetsart University. (in Thai)]
- [21] ภาสพิชญ์ ชูใจ. 2557.การเรียนรู้ร่วมกันสำหรับปัญหาการจำแนกข้อมูลไม่สมดุล.วิทยานิพนธ์ปริญญาโทบริหารธุรกิจ, สาขาวิศวกรรมคอมพิวเตอร์, มหาวิทยาลัยเทคโนโลยีสุรนารี. [PasapitchChujai. 2014. Ensemble Learning for Imbalanced data Classification Problem. Thesis, Computer Engineering Program, Suranaree University of Technology. (in Thai)]
- [22] ภรณ์ยา ปาลวิสุทธิ. 2559. การเพิ่มประสิทธิภาพเทคนิคต้นไม้ตัดสินใจบนชุดข้อมูลที่ไม่สมดุล โดยวิธีการสุ่มเพิ่มตัวอย่างกลุ่มน้อยสำหรับข้อมูล การเป็นโรคติดเชื้อในทางเดินหายใจ. *วารสารเทคโนโลยี*

- สารสนเทศ, 54-63.[ParanyaPalwisut. 2016. Improving Decision Tree Technique in Imbalanced Data Sets Using SMOTE for Internet Addiction Disorder Data. Information Technology Journal, 54-63.(in Thai)]
- [23] Nitesh, V. Kevin, W. Lawrence, O. and Philip, W. 2002. SMOTE: Synthesis Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, 16, 321-357.
- [24] สุรพงศ์ เอื้อวัฒนามงคล. 2559. การทำเหมืองข้อมูล. พิมพ์ครั้งที่ 1, สำนักพิมพ์สถาบันบัณฑิตพัฒนบริหารศาสตร์.กรุงเทพฯ. [SurapongAuwatanamongkol. 2016. Data Mining. 1sted, NIDA Publisher. Bangkok. (in Thai)]
- [25] กันยาพร หาญกล้า. และวีรานันท์ พงศาภักดี. 2556. การเปรียบเทียบตัวแบบลอจิสต์และการใช้ ROC Curve ในการวิเคราะห์ปัจจัยที่ส่งผลต่อระดับค่าจ้าง. *วารสารวิทยาศาสตร์บูรพา*;18 (2556): 26-35.[KanyaphornHankla. and VeeranunPongsapukdee. 2013.A Comparison of Logit Models and Use of ROC Curve in the Analysis of Factors Affecting Wages' Levels. *Burapha Sci. J.* 18(2013): 26-35. (in Thai)]
- [26] Han, J. Kamber, M. and Pei, J. 2012. Data Mining: Concepts and Techniques, 3rd edition, Morgan Kaufmann Publishers, MA, USA.
- [27] Hosmer, D.W. and Lemeshow, S. 2000. Applied Logistic Regression 2nd ed. Wiley, Inc., New York.
- [28] ภิญโญ วรรณสุข. 2540. การประยุกต์ใช้การวิเคราะห์การถดถอยโลจิสติกและการวิเคราะห์อิทธิพลในการศึกษาปัจจัยที่มีอิทธิพลต่อความคาดหวังในการศึกษาต่อของนักเรียนชั้นประถมศึกษาในเขตการศึกษา 11. วิทยานิพนธ์ครุศาสตรมหาบัณฑิต, สาขาวิจัยการศึกษา, จุฬาลงกรณ์มหาวิทยาลัย. [PinyoVannasuk. 1997. The Application of Logistic Regression Analysis and Path Analysis to study Factors affecting lelmaetary school students conyinuting education expectation in educational region eleven. Thesis, Educational Research Program, Chulalongkorn University. (in Thai)]
- [29] Yusuff H., Mohamad N., Ngah U.K. and Yahaya A.S.. Breast cancer analysis using logistic regression. *International Journal of Recent Research and Applied Studies*, 10(1), 14-22.