

ชื่อวิทยานิพนธ์	แบบจำลองทางสถิติของข้อมูลทางการแพทย์และสาธารณสุข: ระยะเวลาที่พักรักษาของผู้ป่วยที่เสียชีวิตในโรงพยาบาล และความรุนแรงของเหตุการณ์ความไม่สงบในภาคใต้ของประเทศไทย
ผู้เขียน	นายวัชรินทร์ โกมลมาลัย
สาขาวิชา	วิธีวิทยาการวิจัย
ปีการศึกษา	2554

### บทคัดย่อ

แบบจำลองและการวิเคราะห์ทางสถิติ ที่นำมาประยุกต์ใช้กับข้อมูลทางการแพทย์และสาธารณสุข โดยวิทยานิพนธ์ฉบับนี้ ได้ใช้ข้อมูลจาก 2 ชุดข้อมูล เพื่อค้นหาแบบจำลองทางสถิติที่เหมาะสม โดยข้อมูลชุดที่ 1 เป็นข้อมูลทางด้านการแพทย์จากสำนักงานหลักประกันสุขภาพแห่งชาติ กระทรวงสาธารณสุข ซึ่งเป็นชุดข้อมูลที่เกี่ยวข้องกับผู้ป่วยที่ดองพักรักษาในโรงพยาบาลในประเทศไทยระหว่างเดือนตุลาคม 2542 ถึงกันยายน 2550 ข้อมูลชุดที่ 2 เป็นข้อมูลทางด้านสาธารณสุขที่เกี่ยวกับเหยื่อผู้เคราะห์ร้าย จากความรุนแรงของเหตุการณ์ความไม่สงบในภาคใต้ของประเทศไทย ซึ่งจัดเก็บโดยศูนย์ประสานงานวิชาการให้ความช่วยเหลือผู้ได้รับผลกระทบจากความไม่สงบจังหวัดชายแดนใต้ (ศวชต.) ตั้งแต่ปี 2547 - 2552 โดยที่แบบจำลองและการประยุกต์ใช้การวิเคราะห์ข้อมูลที่เหมาะสม ได้ถูกนำมาใช้กับข้อมูลประเภทต่างๆ ที่แตกต่างกันดังนี้ ประการที่ 1 สำหรับข้อมูลเชิงตัวเลข ที่มีลักษณะเป็นรายบุคคล ได้ถูกวิเคราะห์ด้วยสถิติการถดถอยเชิงเส้นของลอการิทึม (log-transformed linear regression) โดยปรับประยุกต์ใช้กับข้อมูลที่เป็นศูนย์ ประการที่ 2 การใช้การถดถอยลอจิสติกเชิงพหุ (multiple logistic regression) ซึ่งเป็นวิธีการทางสถิติที่ใช้กันอย่างแพร่หลายในการจำลองรูปแบบความสัมพันธ์ระหว่างตัวแปรตามแบบ 2 ค่า (binary outcome) กับชุดของตัวแปรพยากรณ์ ประการสุดท้ายแบบจำลองเชิงเส้นวางนัยทั่วไป (Generalized linear model: GLMs) โดยมีการแจกแจงแบบปัวซองและทวินามนิเสธ (Poisson and negative binomial) ที่เกิดขึ้นจากจำนวนเหตุการณ์ที่หารด้วยประชากรในกลุ่มเสี่ยง ซึ่งวิธีการทางสถิติสำหรับแบบจำลองรูปแบบนี้ถูกประยุกต์ใช้กับอัตราอุบัติการณ์ โดยที่แบบแผนของความสัมพันธ์ระหว่างตัวแปรตามกับตัวแปรพยากรณ์จะถูกระบุได้อย่างถูกต้องจากรูปแบบจำลองที่เหมาะสม

โดยสรุปการประยุกต์ใช้งานแบบจำลองกับข้อมูลระยะเวลาที่พักรักษาของผู้ป่วย (LOS data) ที่ประกอบด้วยข้อมูลที่มีความเบ้สูง การถดถอยลอจิสติกเชิงพหุ แบบถ่วงค่าน้ำหนักเพื่อปรับค่าความแตกต่าง (weighted sum contrasts) ถูกใช้เพื่อเปรียบเทียบสัดส่วนของระยะเวลาที่พักรักษาที่นานกว่า ระหว่างตัวแปรพยากรณ์แต่ละตัว เมื่อปรับเทียบกับตัวแปรร่วมอื่นๆ (covariates) โดยที่วิธีการนี้จะให้ผลสรุปที่สอดคล้องกับแบบจำลองการถดถอยเชิงเส้นของลอการิทึม แต่การวิเคราะห์การถดถอยลอจิสติกจะได้รับประโยชน์จากการไม่มีข้อตกลงเบื้องต้นเกี่ยวกับการแจกแจงแบบปกติ และรูปแบบความสัมพันธ์เชิงเส้นตรง

ประเด็นต่อไป จะเป็นการการประยุกต์ใช้งานแบบจำลองทวินามนิเสธ และแบบจำลองการถดถอยเชิงเส้นของลอการิทึม เพื่อวิเคราะห์อัตราอุบัติการณ์การบาดเจ็บและตายของชาวไทยมุสลิม จากเหตุการณ์ความไม่สงบในภาคใต้ของประเทศไทย ซึ่งจำแนกตามเพศ กลุ่มอายุ พื้นที่เสี่ยงและปีที่เกิดเหตุการณ์ โดยมีวัตถุประสงค์เพื่อเปรียบเทียบอัตราอุบัติการณ์การบาดเจ็บและตายกับความเสี่ยงที่เกิดขึ้นจากปัจจัยที่ศึกษา โดยปรับลดปัจจัยกวนอื่นๆ แล้ว โดยแต่ละรูปแบบจำลองจะให้สรุปผลที่แตกต่างกัน แต่อย่างไรก็ตามแต่ละรูปแบบจำลองยังคงสามารถระบุถึงพื้นที่ที่มีความเสี่ยงสูง ที่แตกต่างกันไปตามแต่ละช่วงเวลา โดยที่รูปแบบของความเสี่ยงต่อการบาดเจ็บและตายที่เกิดขึ้นนี้จะไม่สามารถพยากรณ์ล่วงหน้าได้ ความเสี่ยงที่แตกต่างกันจำแนกตามเพศและกลุ่มอายุ ยังคงเกิดขึ้นแปรเปลี่ยนไปตามสถานการณ์และจำนวนเหตุการณ์ที่เกิดขึ้น

**Thesis Title** Statistical Modeling of Medical and Public Health Data: Length of Stay of Patients Dying in Hospitals and Terrorism Violence in Southern Thailand

**Author** Mr. Watcharin Komolmalai

**Major Program** Research Methodology

**Academic Year** 2011

### ABSTRACT

The statistical analysis and modeling were applied to medical and public health data. In this thesis, two datasets were used to find the appropriated statistical model: one which using medical dataset from the National Health Security Office (NHSO), the Ministry of Public Health, which information provided to hospitalized patients in Thailand between October 1999 and September 2007. The other public health dataset, terrorism victims in Southern Thailand was recorded by the Deep South Coordination Centre (DSCC) from 2004 to 2009. The appropriate application of statistical analysis and modeling are used in the different types of data. Firstly, for numerical data by individual datasets are used to analyze with log-transformed linear regression modified to handle zero counts. Secondly, multiple logistic regressions that are a statistical method widely used to model the association between a binary outcome and a set of fixed determinants. Finally, generalized linear models (GLMs) by Poisson and negative binomial distributions arise naturally as random counts with population at risk denominators. These statistical methods for the modeling were applied for incidence rates. The association patterns of outcome and determinant variables are identified by fitting the appropriate model.

In conclusion, modeling with application to LOS data that contains outliers, extreme data and skewness, multiple logistic regressions with weighted sum contrasts were used to compare proportions of longer LOS for each factor after adjusting for categorical covariates. These methods provided results consistent with a log-linear regression model. But logistic regression has an advantage of no assumptions related to the distribution, a linear relationship, or equal variances.

Next, the model fitted by negative binomial and log-transformed linear regression models to injury incidence rates of Muslim victims of terrorism violence in southern Thailand, classified by gender, age group, region and year, with the objective of comparing relative risk by these factors, after adjusting for other factors to remove confounding. The models gave different results, but each showed that while specific regions were at higher risk at different times and these patterns could not be easily predicted, risks in different demographic groups remained relatively constant.