

อภินิชา ห่อนบุญเหิม: การคัดเลือกตัวแปรและการประมาณค่าสัมประสิทธิ์แบบเบสเชิง
 ประจักษ์สำหรับตัวแบบ Cox's proportional hazard ที่ข้อมูลมีมิติสูง (EMPIRICAL
 BAYES VARIABLE SELECTION AND ESTIMATION FOR THE COX'S
 PROPORTIONAL HAZARD MODEL WITH HIGH DIMENSIONAL DATA).

อ.ที่ปรึกษาวิทยานิพนธ์หลัก: อ.ดร. วิฐูรา พึ่งพาพงศ์, 78 หน้า.

ในการคัดเลือกตัวแปรและประมาณค่าสัมประสิทธิ์สำหรับตัวแบบ Cox's proportional hazard ที่ข้อมูลมีมิติสูง สามารถวิเคราะห์ได้หลายวิธี ในการศึกษาครั้งนี้เราจะใช้วิธีการคัดเลือกตัวแปรและประมาณค่าสัมประสิทธิ์แบบเบสเชิงประจักษ์ร่วมกับวิธีการทำซ้ำแบบมีเงื่อนไขฐานนิยมและมัธยฐาน ซึ่งเป็นเทคนิคที่สามารถคำนวณได้ง่ายและรวดเร็ว งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาผลกระทบของอัตราส่วนของขนาดตัวอย่างต่อจำนวนตัวแปรอิสระ, ร้อยละของข้อมูลเซ็นเซอร์และค่าสัมประสิทธิ์ถดถอยเริ่มต้นจากค่าอัตราความผิดพลาดในการตรวจจับเชิงบวกและเชิงลบ

ข้อมูลที่ใช้ในการศึกษาคือเวลาในการอยู่รอดที่มีการแจกแจงแบบไวบูลล์ ที่ค่าสัมประสิทธิ์ส่วนใหญ่เป็นศูนย์กำหนดให้ขนาดตัวอย่างเท่ากับ 100 ต่อตัวแปรอิสระ 300, 500 และ 1,000 ที่ร้อยละของข้อมูลเซ็นเซอร์ 10%, 50% และ 70% พบว่า i) อัตราความผิดพลาดในการตรวจจับเชิงบวกและเชิงลบจะต่ำเมื่อระดับของข้อมูลเซ็นเซอร์ต่ำ ii) อัตราความผิดพลาดในการตรวจจับเชิงบวกและเชิงลบจะต่ำเมื่อจำนวนตัวแปรอิสระน้อย และ iii) อัตราความผิดพลาดในการตรวจจับเชิงบวกและเชิงลบจะต่ำเมื่อค่าสัมประสิทธิ์ถดถอยเริ่มต้นเป็นค่าจริง vi) วิธีแบบเบสเชิงประจักษ์ดีกว่าวิธีแบบ penalized lasso

ภาควิชา.....สถิติ.....
 สาขาวิชา.....สถิติ.....
 ปีการศึกษา.....2555.....

ลายมือชื่อนิสิต อภินิชา ห่อนบุญเหิม.....
 ลายมือชื่อ อ.ที่ปรึกษาวิทยานิพนธ์หลัก.....
 ๖๓ ๗

5481721926: MAJOR STATISTICS

KEYWORDS: EMPIRICAL BAYES/COX'S PROPORTIONAL HAZARD MODEL/HIGH
DIMENSIONAL DATA/ SPARSE VARIABLE

ONNICA HONBOONHERM: EMPIRICAL BAYES VARIABLE SELECTION AND
ESTIMATION FOR THE COX'S PROPORTIONAL HAZARD MODEL WITH HIGH
DIMENSIONAL DATA. ADVISOR: VITARA PUNGPAONG, Ph.D., 78 pp.

Cox's proportional hazard model with high-dimensional data can analyses in several ways. In this study we will use empirical Bayes variable selection methods combined with iterated conditional modes/medians (ICM/M) algorithm which is empirically faster and easy to implement. The objective of this dissertation is to study the effects from the ratio of sample size to the number of independent variables, the percentages of censored data and the value of initial coefficient by comparing false positive and false negative rate.

The data in this study is survival times with Weibull distribution. Simulate sparse data with 100 sample size and 300, 500 and 1000 independent variables. The levels of percentages of censored data are 10%, 50% and 70%. Based on the false positive and false negative rate, the findings are following: i) the false positive and false negative rate will decrease as low percentage of censored data, ii) false positive and false negative rate will decrease as the number of variables is small and iii) false positive and false negative will decrease as the initial value of coefficients is true coefficients iv) empirical Bayes method is better than the Lasso method.

Department:.....Statistics.....

Student's Signature.....อนันต์ หอมนาน.....

Field of Study:.....Statistics.....

Advisor's Signature.....เจน.....

Academic Year:.....2012.....