ชื่อผู้วิจัย รองศาสตราจารย์มนตรี พิริยะกุล

ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยรามคำแหง

ปีที่ทำวิจัย พ.ศ. 2547-2548

173274

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อประเมินผลและเปรียบเทียบพฤติกรรมการทำงานของตัวแบบทดแทนข้อมูลสูญหาย (missing data technique, MDT) จำนวน18 วิธีทั้งในกรณีเมื่อใช้กับข้อมูลจริงและเมื่อใช้กับข้อมูลเลียนแบบ โดยศึกษาเฉพาะในสถานการณ์ที่มีข้อมูลสูญหายไปโดยสุ่ม (missing at random, MAR) กำหนดให้มีสัดส่วนหน่วยสำรวจไม่สมบูรณ์ (m_c) รวม 4 ระดับคือ 5 % 10 % 15 % และ 20 % กำหนดให้มีสัดส่วนข้อถามที่ไม่ได้รับการตอบสนอง (m_i) รวม 3 ระดับคือ 10 % 20 % และ 30 % กำหนดให้มีขนาดตัวอย่าง (n) รวม 3 ขนาดคือ 100 200 และ 500 แต่ละเงื่อนไขการทดลองทำการทดลอง 100 ซ้ำ เฉพาะในกรณีของการทดลองมอนติคาร์โลจะเพิ่มเงื่อนไขเกี่ยวกับระดับของภาวะร่วมเส้นตรงพหุรวม 5 ระดับลงในเงื่อนไขการทดลองด้วย

การทดลองมอนติคาร์โลกระทำกับข้อมูลเลียนแบบจำนวน 18,000 กลุ่มตัวอย่างและการทดลองวิเคราะห์เชิงประจักษ์กระทำกับข้อมูลจริงจากแฟ้มข้อมูลเรื่องการสำรวจเกี่ยวกับอนามัยและสวัสดิการ พ.ศ. 2544 สำรวจโดยสำนักงานสถิติแห่งชาติที่มีข้อมูลทั้งสิ้น 118,285 ระเบียน คัดเอามาเฉพาะระเบียนสมบูรณ์ที่สอดคล้องกับข้อถามที่เลือกมาทดลองจำนวน 15 ข้อถามรวม 83,641 ระเบียนจำนวน 3,600 กลุ่มตัวอย่าง เพื่อใช้เปรียบเทียบคุณภาพของการทดแทนข้อมูลในระหว่างวิธีทดแทนข้อมูล 18 วิธี กำหนดให้ใช้สถิติตัดสินใจ 4 ตัวคือ

173274

ค่าเฉลี่ยของ MSE ค่าเฉลี่ยของ bias และค่าเฉลี่ยของสหสัมพันธ์ (r) และค่าเฉลี่ยของระดับความเชื่อถือได้ (Cronbach's α) แต่ในขั้นปฏิบัติมิได้ใช้ค่าเฉลี่ยของระดับความเชื่อถือได้ เนื่องจากค่าที่ได้จากการวิเคราะห์ไม่ชี้ให้เห็นความแตกต่างระหว่าง MDT

ผลการศึกษาพบว่าวิธีทดแทนข้อมูลสูญหายที่ดีจำแนก ได้เป็น 2 กลุ่มคือ

กลุ่มที่ 1 กรณีเมื่อข้อมูลวัดค่าตามมาตรานามบัญญัติจะประกอบด้วยวิธี IMS วิธี HDD-IMS และวิธี HDD-IMS-Z ตามลำดับ

กลุ่มที่ 2 กรณีข้อมูลวัดค่าตามมาตราอันดับแบบลิกเกอร์ทจะประกอบด้วยวิธี PMS วิธี HDD-PMS และวิธี HDD-PMS-Z ตามลำดับ

และพบเพิ่มเติมว่าวิธี HDR-IMS วิธี HDR-IMS-Z วิธี HDR-PMS และวิธี HDR-PMS-Z เป็นวิธีที่ดีถ้าพิจารณาในด้านของความง่ายในการเลือกหน่วยให้ข้อมูล (donor)

การพิจารณาเลือกตัวแบบทดแทนจึงขึ้นอยู่กับผู้วิจัย ถ้าผู้วิจัยต้องการความสะดวก ใช้ง่าย เข้าใจง่าย ให้เลือกใช้วิธี IMS หรือวิธี PMS ขึ้นอยู่กับมาตราที่ใช้วัดค่าตัวแปร ถ้าต้องการความแม่นยำให้ใช้วิธีผสมคือวิธีทดแทนข้อมูลแบบ IMS หรือ PMS ผสมกับวิธี HDD แต่โดยข้อเท็จจริงแล้ววิธีทดแทนข้อมูลทั้ง 3 วิธีในกลุ่มเดียวกันมีคุณภาพไม่แตกต่างกัน รวมทั้งไม่ผูกพันกับขนาดตัวอย่าง อัตราการสูญหายของข้อมูล และระดับภาวะร่วมเส้นตรงพหุ

This study is aimed at evaluation and comparison of imputation behavior, under missing at random (MAR) pattern, among 18 different missing data techniques (MDT) where various designs were planned - i.e. combination of 4 m_c (missing case of 5 %, 10 %, 15 % and 20 %), 3 m_i (missing item of 10 %, 20 % and 30 %) and 3 sample size (n = 100, 200, 500) with one hundred replications each. In addition, 5 levels of multicollinearity were added if Monte Carlo experiment was to investigated. An empirical analysis depended its data on file of national health and welfare survey conducted by NSO in 2001. There were 18,000 samples employed in simulation study while 3,600 samples did in empirical analysis.

Experiments in Monte Carlo fashion with Likert type data and real nominal data were processed separately and average values of MSE, bias, Pearson product moment and Cronbach's α were extracted and plotted. From these indices, except α which discriminated nothing, it was found that there were 2 groups of MDT's judged to be optimal with no obligations of sample size, percentage of missing cases, percentage of missing items and multicollinearity level; they are IMS, HDD-IMS and HDD-IMS-Z for real nominal data and PMS, HDD-PMS and HDD-PMS-Z for LIKRT scaled data. Among optimal MDT's in either group, any MDT can be chosen to be used arbitrarily since they are equally better.

However, for ease of use and less computation arrangement, IMS and PMS are recommended, but for more accurate and general usage, HDD-IMS and HDD-IMS-Z or HDD-PMS and HDD-PMS-Z are suggested.