

# ตัวแบบการทดแทนข้อมูลในการวิจัยทางสังคมศาสตร์

## : การจำลองแบบกรณีตัวอย่างง่าย

บทคัดย่อ

รองศาสตราจารย์ ดร. มนตรี พิริยะกุล<sup>๑</sup>

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อประเมินผลและเปรียบเทียบพฤติกรรมการทำงานของตัวแบบทดแทนข้อมูลสูญหาย (missing data technique, MDT) จำนวน 18 วิธีทั้งในกรณีเมื่อใช้กับข้อมูลจริงและเมื่อใช้กับข้อมูลเลียนแบบ โดยศึกษาเฉพาะในสถานการณ์ที่มีข้อมูลสูญหายไปโดยสุ่ม (missing at random, MAR) กำหนดให้มีสัดส่วนหน่วยสำรวจไม่สมบูรณ์ ( $m_1$ ) รวม 4 ระดับคือ 5 % 10 % 15 % และ 20 % กำหนดให้มีสัดส่วนข้อถามที่ไม่ได้รับการตอบสนอง ( $m_2$ ) รวม 3 ระดับคือ 10 % 20 % และ 30 % กำหนดให้มีขนาดตัวอย่าง ( $n$ ) รวม 3 ขนาดคือ 100 200 และ 500 แต่ละเงื่อนไขการทดลองทำการทดลอง 100 ครั้ง เฉพาะในกรณีของการทดลองมอนติคาร์โลจะเพิ่มเงื่อนไขเกี่ยวกับระดับของภาวะร่วมเส้นตรงพหุรวม 5 ระดับลงในเงื่อนไขการทดลองด้วย

การทดลองมอนติคาร์โลกระทำกับข้อมูลเลียนแบบจำนวน 18,000 กลุ่มตัวอย่าง และการทดลองวิเคราะห์เชิงประจักษ์ที่กระทำกับข้อมูลจริงจากแฟ้มข้อมูลเรื่องการสำรวจเกี่ยวกับอนามัยและสวัสดิการ พ.ศ. 2544 สำรวจโดยสำนักงานสถิติแห่งชาติที่มีข้อมูลทั้งสิ้น 118,285 ระเบียบ คัดเอามาเฉพาะระเบียบสมบูรณที่สอดคล้องกับข้อถามที่เลือกมาทดลองจำนวน 15 ข้อถามรวม 83,641 ระเบียบจำนวน 3,600 กลุ่มตัวอย่าง เพื่อใช้เปรียบเทียบคุณภาพการทดแทนข้อมูลในระหว่างวิธีทดแทนข้อมูล 18 วิธี กำหนดให้ใช้สถิติตัดสินใจ 4 ตัว คือ ค่าเฉลี่ยของ MSE ค่าเฉลี่ยของ bias และค่าเฉลี่ยของสหสัมพันธ์ ( $r$ ) แต่มิได้ใช้ค่าเฉลี่ยของระดับความเชื่อถือได้ (Cronbach's  $\alpha$ ) เป็นเครื่องมือเนื่องจากไม่ชี้ให้เห็นความแตกต่างระหว่าง MDT

ผลการศึกษาพบว่าวิธีทดแทนข้อมูลสูญหายที่ดีที่สุดจำแนก ได้เป็น 2 กลุ่มคือ

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

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

วิธี IMS คือวิธีแทนที่ค่าของตัวแปรที่มีข้อมูลสูญหายด้วยค่าเฉลี่ยของตัวแปรเดียวกันจากระเบียบอื่น

วิธี HDD-IMS คือวิธีแทนที่ค่าของตัวแปรที่มีข้อมูลสูญหายด้วยค่าของตัวแปรเดียวกันจากระเบียบอื่นที่มีสมบัติทางประชากรคล้ายคลึงกันกับระเบียบที่มีข้อมูลสูญหายมากที่สุด ความคล้ายคลึง

คำสำคัญ : มอนติคาร์โล, การวิเคราะห์เชิงประจักษ์, MDT, Hot-Deck

รองศาสตราจารย์มนตรี พิริยะกุล : mpiriyakul@yahoo.com

<sup>๑</sup> ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยรามคำแหง

(similarity) วัดด้วยค่าระยะทาง ทั้งนี้ก่อนที่จะคำนวณค่าความคล้ายคลึงให้แทนที่ข้อมูลที่สูญหายตามวิธี IMS เสียก่อน

วิธี HDD-IMS-Z คือวิธีแทนที่ค่าของตัวแปรที่มีข้อมูลสูญหายด้วยค่าของตัวแปรเดียวกันจากระเบียนอื่นที่มีสมบัติทางประชากรคล้ายคลึงกับระเบียบที่มีข้อมูลสูญหายมากที่สุด ก่อนที่จะคำนวณค่าความคล้ายคลึงให้แทนที่ข้อมูลที่สูญหายตามวิธี IMS และแปลงค่าทุกตัวแปรให้เป็นคะแนนมาตรฐานเสียก่อน

วิธี PMS คือวิธีแทนที่ค่าของตัวแปรที่มีข้อมูลสูญหายด้วยค่าเฉลี่ยของทุกตัวแปรจากระเบียนเดียวกัน

วิธี HDD-PMS คือวิธีแทนที่ค่าของตัวแปรที่มีข้อมูลสูญหายด้วยค่าของตัวแปรเดียวกันจากระเบียนอื่นที่มีสมบัติทางประชากรคล้ายคลึงกับระเบียบที่มีข้อมูลสูญหายมากที่สุด ความคล้ายคลึงวัดด้วยค่าระยะทาง ทั้งนี้ก่อนที่จะคำนวณค่าความคล้ายคลึงให้แทนที่ข้อมูลที่สูญหายตามวิธี PMS เสียก่อน

วิธี HDD-PMS-Z คือวิธีแทนที่ค่าของตัวแปรที่มีข้อมูลสูญหายด้วยค่าของตัวแปรเดียวกันจากระเบียนอื่นที่มีสมบัติทางประชากรคล้ายคลึงกับระเบียบที่มีข้อมูลสูญหายมากที่สุด ก่อนที่จะคำนวณค่าความคล้ายคลึงให้แทนที่ข้อมูลที่สูญหายตามวิธี PMS และแปลงค่าทุกตัวแปรให้เป็นคะแนนมาตรฐานเสียก่อน

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

การพิจารณาเลือกตัวแบบสำหรับสร้างข้อมูลทดแทนจึงขึ้นอยู่กับผู้วิจัย ถ้าผู้วิจัยต้องการความสะดวก ใช้ง่าย เข้าใจง่าย ให้เลือกใช้วิธี IMS หรือวิธี PMS ขึ้นอยู่กับมาตราที่ใช้วัดค่าตัวแปร ถ้าต้องการความแม่นยำให้ใช้วิธีผสมคือวิธีทดแทนข้อมูลแบบ IMS หรือ PMS ผสมกับวิธี HDD แต่โดยข้อเท็จจริงแล้ววิธีทดแทนข้อมูลทั้ง 3 วิธีในกลุ่มเดียวกันมีคุณภาพไม่แตกต่างกัน รวมทั้งไม่ผูกพันกับขนาดตัวอย่าง อัตราการสูญหายของข้อมูลและระดับภาวะร่วมเส้นตรงพหุ

# Missing Data Replacement Models in Social Science Research : Simulation Study of Simple Model

Montree Piriyaikul<sup>b</sup>

## Abstract

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 investigated. An empirical analysis depended its data on file of National Health and Welfare Survey conducted by NSO, Thailand 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 simulated Likert type data and in empirical analysis style with real nominal data were processed separately whereas average values of MSE, bias, Pearson product moment and Cronbach's  $\alpha$  were extracted and plotted. From 3 of these 4 evidences, except Cronbach's  $\alpha$  which discriminated nothing. 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 simulated LIKRT scaled data. Among optimal MDT's in either group, they can be chosen to be used arbitrarily since the different among their average value of MSE or bias or correlation coefficient are subjectively small.

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.

**Keyword** : Monte Carlo, Empirical analysis, MDT, Hot-Deck

**Associated professor Montree Piriyaikul:** mpiriyakul@yahoo.com

---

<sup>b</sup> Department of Statistics, Ramkhamhaeng University

## บทนำ

ปัญหาข้อมูลสูญหายเป็นเรื่องที่นักวิจัยต้องพบเป็นประจำ (Heeringa, 2000; DeSilvio, 1999; Wesenbaker, 1998; Samuhel, 1983) เป็นเรื่องที่ต้องเกิดขึ้นที่นักวิจัยจะต้องเผชิญ แม้ว่าจะได้ควบคุมการสำรวจหรือการทดลองไว้แล้วเป็นอย่างดีก็ตาม (Huisman, 1999) ซึ่งนักวิจัยทางสังคมศาสตร์กำลังให้ความสนใจเพิ่มขึ้นเรื่อย ๆ (Adam, 2001) ปัญหาข้อมูลสูญหายอาจไม่ถือว่าเป็นปัญหาที่รุนแรงหรืออาจถือว่าเป็นเรื่องเล็กน้อยถ้าการวิเคราะห์ข้อมูลกระทำด้วยสถิติที่วิเคราะห์ข้อมูลรายตัวแปร เช่นร้อยละหรือค่าเฉลี่ยหรือสถิติพรรณนาอื่น แต่ถ้าการวิเคราะห์ข้อมูลนั้นจำเป็นต้องใช้วิธีวิเคราะห์พหุ เช่นการวิเคราะห์ความถดถอยพหุ การวิเคราะห์เส้นทาง การวิเคราะห์ปัจจัย การวิเคราะห์กลุ่ม และการวิเคราะห์การจำแนกพหุ ในกรณีนี้การสูญหายของข้อมูลจะมีผลกระทบต่อที่รุนแรง เพราะถ้าพบว่าหน่วยวิเคราะห์ใดมีตัวแปรใดขาดข้อมูลไปแม้เพียงตัวแปรเดียวก็จะตัดหน่วยวิเคราะห์นั้นทิ้งทั้งหน่วย โดยไม่สนใจว่าจะยังมีตัวแปรอื่นอีกมากที่มีข้อมูลครบถ้วนหรือไม่ (Heeringa, 2000; Roth, 1994) ซึ่งการตัดข้อมูลทิ้งเป็นวิธีปริยาย (default) ของการวิเคราะห์พหุในโปรแกรม SPSS หรือโปรแกรมสถิติทั่วไป จากการศึกษาโดยวิธีทดลองพบว่าถ้าตัวแปรแต่ละตัวมีข้อมูลหายไปโดยสุ่มเพียง 10% จะมีผลให้ต้องตัดหน่วยวิเคราะห์ทิ้งถึง 59 % (Kim & Curry, 1977 อ้างถึงใน Roth, 1995) เห็นได้ว่าเป็นความสูญเสียในอัตราที่สูงมาก การวิเคราะห์ข้อมูลจากเฉพาะข้อมูลที่เหลืออยู่ภายหลังจากที่ได้ตัดทิ้งค่าสังเกตที่ไม่สมบูรณ์ไปแล้ว ผลการวิเคราะห์จะเอนเอียง (bias) และไม่ถูกต้อง (Wang, 2000)

การไม่ตอบแบบสำรวจเป็นเรื่องทางสังคม อีกทั้งยังเป็นเรื่องที่เกี่ยวข้องกับบุคลิกภาพและจิตวิทยา แม้จะมีกลไกทางสถิติช่วยลดอัตราการไม่ตอบแบบสำรวจลง แต่การไม่ตอบคำถามจะยังคงปรากฏอยู่เพราะสาเหตุหลักเป็นเรื่องทางสังคม (Ruben, 1992) เป็นเรื่องของการแลกเปลี่ยนทางสังคมที่หากผู้ตอบไม่ได้รับประโยชน์ที่อาจเป็นด้านวัตถุหรือด้านจิตใจเป็นการตอบแทนแลกเปลี่ยนกับการให้ข้อมูลก็จะไม่ให้ข้อมูล และยังเป็นเรื่องปฏิสัมพันธ์สัญลักษณ์ (Murata, 2001) เช่น ถ้อยคำ วลี ประโยค และการสื่อสารที่ใช้ในกระบวนการถามตอบสื่อความหมายไม่ชัดเจน สื่อไม่ตรงกันตลอดกับผู้ตอบทุกคน ในบริบทเดียวกัน ยกเว้นไป มีผังการจัดวางคำถามที่ไม่เหมาะสม ฯลฯ ที่ล้วนส่งผลต่อปฏิสัมพันธ์ข้อมูลก็จะขาดหาย เกิดการสูญเสีย ขาดสารสนเทศที่เพียงพอแก่การสรุปผล ตัวแบบทางสถิติสำหรับทดแทนข้อมูลจึงเป็นสิ่งจำเป็นที่เข้ามาเสริมเพื่อแทนที่ข้อมูลที่ขาดหาย ทำให้ผลการศึกษาน่าเชื่อถือขึ้นมองเห็นภาพชัดเจนขึ้น มีความสมบูรณ์มากขึ้น สรุปผลได้ตรง ไม่เอนเอียง

ค่าประมาณของข้อมูลที่สูญหายแม้จะเป็นของเทียมแต่ก็สามารถแทนที่ค่าสังเกตจริงได้ ความจริงแล้วแม้แต่ผู้ตอบแบบสอบถามเองก็ยังให้ข้อมูลที่ไม่แน่นอน เชื่อไม่ได้ว่าจะมีความคิดเห็นเช่นนั้นหรือมีข้อมูลเป็นปริมาณเท่านี้ตลอดไป เพราะเมื่อเวลาเปลี่ยนไปเพียงระยะเวลาหนึ่งผู้ตอบอาจเปลี่ยนใจหรือเปลี่ยนคำตอบได้ จากการศึกษาโดยติดตามถามซ้ำหน่วยเดิมที่ไม่ยอมตอบคำถามบางคำถามและถามซ้ำคำถามเดิมที่ได้ตอบมาแล้ว ปรากฏว่า 95% ของผู้ตอบเปลี่ยนคำตอบ (Huisman, Kron, & Van Sonderen, 1998) จึงเห็นได้ว่าแม้แต่คำตอบจริงที่ได้จากผู้ตอบเองก็ยังเปลี่ยนแปลงได้ จึงมีสภาพคล้ายค่าทดแทน การแย้งว่าค่าทดแทนเป็นเพียงค่าประมาณมิใช่ค่าจริงจึงต้องพิจารณาความจริงข้อนี้ แม้การ

ประมาณค่าสังเกตเหล่านี้ด้วยค่าทดแทนที่มีใช้ค่าจริงแต่ก็เป็นค่าที่ได้ผ่านการทดลองซ้ำ ๆ แล้วว่ามีความแม่นยำสูงอย่างน่าพอใจ การทดแทนข้อมูลที่สูญหายด้วยตัวแบบที่เหมาะสมดังกล่าวจึงน่าจะยอมรับได้ ข้อที่ต้องคำนึงถึงเสมอคือ แม้ว่าข้อมูลทดแทนจะเป็นเพียงค่าประมาณแต่ก็ยิ่งดีกว่าทั้งค่าสังเกตที่สูญหายนั้นไปเฉย ๆ เพราะจะเกิดความเสียหายมากกว่า มิเช่นนั้นนักวิจัยจะต้องใช้วิธีติดตามสำรวจซ้ำ

การไม่ตอบคำถามจะนำไปสู่การสูญหายของข้อมูลบางรายการ และการที่เรารู้เฉพาะชื่อข้อมูลทั้ง ๆ ที่มีข้อมูลสูญหายจะนำไปสู่ความยุ่งยากทั้งในทางทฤษฎีและในทางปฏิบัติ

1. ถ้ารายการข้อมูลที่สูญหายไปนั้นไม่ได้มีค่าใกล้เคียงกับข้อมูลของตัวแปรเดียวกันที่เหลืออยู่ แต่กลับมีค่าผิดแผกแตกต่างไปเป็นอย่างมาก การวิเคราะห์ข้อมูลจากข้อมูลเท่าที่มีอยู่จะได้คำตอบที่ผิดจากความจริง นำสู่การสรุปผลการวิจัยผิดพลาด อนึ่ง เมื่อมีข้อมูลสูญหายเกิดขึ้นจะพบว่าค่าสถิติที่ใช้วัดแนวโน้มสู่ส่วนกลางมีความเอนเอียง (bias) ที่อาจสูงขึ้นหรือต่ำลงก็ได้ (DeSilvio, 1999) คือผลการศึกษามีได้ชี้ไปที่ความจริงของปรากฏการณ์นั้น (Rearden, 1991; Kim, 2000; Belin, 1998; DeSilvio, 1999) การแทนที่ข้อมูลที่สูญหายด้วยค่าทดแทนที่เหมาะสมจะลดความเอนเอียงลงได้ (Kwang, 2000) ค่าเฉลี่ยที่เอนเอียงหมายถึงค่าเฉลี่ยที่ชี้ไปที่ใดที่หนึ่งที่มีใช้ค่าจริง ค่าสถิติจึงมีได้แม่นยำตามที่คาดหวัง ทั้งยังเป็นค่าที่ไม่ถูกต้อง (Joen, 1998; Ibrahim, 1998; DeSilvio, 1999; Wang, 2000)

นอกจากนี้ ข้อมูลสูญหายยังทำให้ความผันแปรและส่วนเบี่ยงเบนมาตรฐานของข้อมูลผิดไปจากความเป็นจริง อาจมีค่าต่ำเกินไปหรือว่าสูงเกินไปได้ขึ้นอยู่กับตำแหน่งที่ข้อมูลหายไป (Roth, 1995)

ค่าความผันแปรและส่วนเบี่ยงเบนมาตรฐานมีอิทธิพลอย่างยิ่งต่อการวิจัยทางสังคม ค่าส่วนเบี่ยงเบนมาตรฐานที่ต่ำกว่าแสดงให้เห็นเราทราบข้อมูลมีความหลากหลายน้อย ผิดแผกกันไม่มาก เช่น ส่วนเบี่ยงเบนมาตรฐานของรายได้มีค่าต่ำแสดงว่าประชากรเป้าหมายมีฐานะทางเศรษฐกิจไม่ต่างกัน รายจนวนต่างกันไม่มาก สังคมน่าจะอบอุ่นราบรื่นเพราะไม่ค่อยมีปัญหาชนชั้น แต่ที่สรุปเช่นนี้เพราะมีข้อมูลส่วนหนึ่งหายไป ที่จริงอาจแล้วอาจต้องสรุปผลเป็นอย่างอื่นถ้ามีข้อมูลครบถ้วนสมบูรณ์

ค่าเฉลี่ยและค่าความผันแปรที่ผิดพลาดมิได้ส่งผลเพียงเท่านั้น แต่ยังส่งผลให้สถิติ  $t$  สถิติ  $F$  และสหสัมพันธ์ ( $r$ ) มีค่าผิดพลาด รวมทั้งค่าสถิติวัดคุณภาพของมาตรวัดผิดพลาดจนเราไม่แน่ใจว่ามาตรวัดที่นำไปใช้แล้วนั้นวัดในสิ่งที่ต้องการวัดได้จริงหรือไม่

2. ค่าประมาณที่ดีต้องอาศัยข้อมูลของตัวแปรที่สนใจที่มากพอ และต้องมีความหลากหลายหรือผันแปรครบทุกความเป็นไปได้ เมื่อมีข้อมูลสูญหายทำให้เราสูญเสียสารสนเทศที่จำเป็น ค่าประมาณที่ขาดสารสนเทศที่เพียงพอจะไม่สามารถเป็นตัวแทนที่เหมาะสมของสิ่งที่ต้องการประมาณค่าได้

3. สถิติที่ใช้วิเคราะห์นั้นเราได้ออกแบบเอาไว้สำหรับวิเคราะห์กับข้อมูลที่ต้องได้ข้อมูลมาครบถ้วนตามขนาดตัวอย่างที่ได้กำหนดไว้ก่อนการสำรวจ ถ้าข้อมูลสูญหายไปสภาพสุ่มย่อมเสียไปกลายเป็นสภาพสุ่มบางส่วนบางที่ ตัวสถิติที่เหมาะสมสำหรับกรณีเช่นนั้นจะต้องเป็นอย่างอื่น

4. ข้อมูลสูญหายไปทำให้ตัวสถิติที่ใช้สำหรับงานทดสอบสมมุติฐานลดความสามารถในการตัดสินใจลงคือลดอำนาจการทดสอบลง (Witta, 2000; Roth, 1994)

โดยปกติอำนาจการทดสอบสูงจะเกิดขึ้นได้เพราะเราใช้ตัวอย่างขนาดใหญ่ จากการทดลองพบว่าเมื่อกำหนดให้ข้อมูลของตัวแปรหายไปโดยสุ่มเพียง 2% จะมีผลให้ต้องตัดหน่วยสำรวจทิ้งไม่ถึง 18.3% ของตัวอย่างทั้งหมด เห็นได้ว่าแม้จะเป็นการสูญหายเพียงเล็กน้อย คือผู้ตอบบางคนเท่านั้นที่ไม่

ตอบคำถามหรือเราไม่ได้รับข้อมูลบางรายการเท่านั้น แต่กลับมีผลเสียหายอย่างมาก สมมติว่าเรา กำหนดตัวอย่างไว้ 370 หน่วยซึ่งเพียงพอที่จะใช้ประมาณค่าสิ่งที่ต้องการ ด้วยความเชื่อถือได้และระดับ ความถูกต้องตรงตามกำหนด ปรากฏว่าข้อมูลสูญหายไปเพียง 2% ทำให้ต้องตัดหน่วยสำรวจทิ้งถึง 68 หน่วย (Roth, 1995)

5. ข้อมูลสูญหายทำให้ค่าสหสัมพันธ์ ( $r$ ) ที่คำนวณมาจากกลุ่มตัวอย่างคาดคะเนความสัมพันธ์ จริงผิดพลาด โดยความผิดพลาดจะปรากฏในลักษณะของการระบุนความสัมพันธ์ต่ำกว่าที่ควรจะเป็น (downward bias) สาเหตุที่สำคัญคือการสูญหายของข้อมูลทำให้ความผันแปรของตัวแปรเปลี่ยนไป ระดับ ความสัมพันธ์กับตัวแปรอื่นจึงเปลี่ยนแปลงตาม (Kim & Curry, 1997; Mahotra, 1987 อ้างใน Roth, 1994)

6. ในการวัดคุณภาพของมาตรวัด เช่น แบบสอบถาม แบบประเมินต่าง ๆ นั้นเราต้องการมาตร วัดที่สามารถวัดได้ถูกต้อง คือ วัดได้ในสิ่งที่ต้องการวัดและผลการวัดต้องเชื่อถือได้ การสูญหายของข้อมูล ทำให้ทั้งค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานมีค่าเอนเอียงทำให้สถิติวัดคุณภาพแบบทดสอบ ได้รับความ กระทบกระเทือน แบบทดสอบที่ผ่านการวัดไปขาดความเชื่อถือได้

การแก้ปัญหาข้อมูลสูญหายมีหลายวิธี วิธีที่ดีที่สุดควรเป็นวิธีที่ใช้ง่ายที่สุด ไม่เป็นภาระแก่นักวิจัย ที่ไม่สั่นหวั่นวิธีการทางสถิติ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อประเมินผลพฤติกรรมการทำงานของตัวแบบทดแทน (missing data technique, MDT) ในสถานการณ์ที่ข้อมูลมีการสูญหายไปโดยสุ่ม (missing at random, MAR) คือข้อมูลที่ สูญหายเกี่ยวข้องกับลักษณะทางประชากร เช่น อาชีพ รายได้ การศึกษาของผู้ให้ข้อมูล ซึ่งจะได้นำมา ทดลองเพื่อศึกษาเปรียบเทียบกันรวม 18 วิธีโดยใช้ทั้งข้อมูลจริงและข้อมูลเลียนแบบ แต่จำกัดขอบเขต การศึกษาไว้เพียงข้อมูลที่วัดค่าในมาตรฐานอันดับตามมาตรวัดแบบลิเกอ์ทและมาตรฐานนามบัญญัติ และการ สูญหายของข้อมูลเกิดขึ้นกับเฉพาะตัวแปรตามเท่านั้น

### **ตัวแบบการทดแทนข้อมูล**

ตัวแบบที่น่าเสนอนี้ส่วนหนึ่งเป็นวิธีทดแทนข้อมูลสูญหายโดยการแทนที่ข้อมูลที่สูญหายด้วย ค่าเฉลี่ย (Downey & King, 1998; Huisman, Krol & Sonderen, 1998) วิธีที่เหลือคือวิธี hot-deck และวิธี ICS

วิธี hot-deck หรือเรียกอีกอย่างหนึ่งว่าวิธี neighborhood next door หรือ nearest neighbor หรือ closest fit หรือ nearby household (Saurabh, 2003) เป็นวิธีทดแทนข้อมูลที่สูญหายด้วยข้อมูลของผู้ให้ ข้อมูล (donor) อาจโดยวิธีกำหนดคุณสมบัติร่วมของหน่วยให้/รับข้อมูล (categorization) หรือโดยการหา ระยะทาง (euclidian distance) เป็นวิธีที่สำนักงานสำมะโนประชากรของสหรัฐอเมริกา (US Census Bureau) สำนักงานสำมะโนประชากรของแคนาดาและสำนักงานสำมะโนประชากรแห่งสหราชอาณาจักรใช้สร้างข้อมูล ทดแทน (Roth & Switzer, 1995; Paullin & Ferraro, 1994; Hirsch & Schumacher, 2001; Dinardo & Lemieux, 2001; Brunel, 2000; Geoffrey & Ferrado, 1994) สารสังเขปของวิธีทดแทนข้อมูลจำนวน 18 วิธีปรากฏดังนี้

1. **วิธี RDS (Random Draw Substitution)** วิธี RDS คือวิธีทดแทนข้อมูลที่สูญหายโดยเลือกข้อมูล มาโดยสุ่มจากช่วงที่เป็นไปได้ของข้อมูลรายการนั้น พบว่าวิธีนี้เป็นวิธีที่มีคุณภาพต่ำที่สุด (Huisman, 1997)

2. **วิธี HNC (hot-deck next case)** วิธี HNC คือวิธีที่แทนที่ข้อมูลที่สูญหายโดยใช้ข้อมูลจากหน่วยสำรวจที่สมบูรณ์หน่วยแรกที่อยู่ถัดลงไป วิธีนี้มีคุณภาพต่ำแต่โดยทั่วไปจะมีคุณภาพดีกว่าวิธี RDS เล็กน้อย (Huisman, 1997)

3. **วิธี IMS (item mean substitution)** วิธี IMS คือวิธีที่ทดแทนข้อมูลที่เกิดการนำเอาค่าเฉลี่ยจากข้อมูลที่มีอยู่ของรายการข้อมูลที่มีข้อมูลสูญหาย (missing item) มาใช้เป็นข้อมูลทดแทนให้แก่หน่วยสำรวจที่ไม่ตอบคำถาม (ตัวแปร) นั้น

4. **วิธี PMS (Person Mean Substitution)** วิธี PMS คือวิธีที่เราแทนที่ข้อมูลสูญหายด้วยค่าเฉลี่ยคำตอบของหน่วยสำรวจที่ได้ตอบแบบสำรวจนั้น จากการศึกษาพบว่าวิธี PMS เป็นวิธีที่มีคุณภาพสูงกว่าวิธี IMS (Huisman, 1997)

5. **วิธี CIMS (corrected item mean substitution)** เป็นวิธีที่นำเอาทั้งผลกระทบจากบุคคลและผลกระทบจากคำถามมารวมเป็นตัวถ่วงน้ำหนักเพื่อใช้สร้างข้อมูลทดแทน ผลกระทบจากบุคคลคือยอดรวมคำตอบของบุคคล ถ้าผู้ตอบรายใดตอบคำถามมากกว่าคำถามกว่า น้ำหนักของบุคคลผู้นั้นอาจสูงกว่าน้ำหนักของผู้อื่นที่ตอบคำถามน้อยกว่า ถ้าตัวแปรใดมีข้อมูลสูญหายน้อยกว่าตัวแปรนั้นย่อมมีน้ำหนักมากกว่าตัวแปรอื่นที่มีข้อมูลสูญหายมากกว่าวิธี พบว่า CIMS เป็นวิธีที่ดีที่สุด (Huisman, 1997)

$$CIM_{vi} = w_v \bar{x}_i^{(i)} = \left( \frac{\sum_{i \in obs(v)} x_{vi}}{\sum_{i \in obs(v)} \bar{x}_i^{(i)}} \right) \bar{x}_i^{(i)} \quad ; v = 1, 2, \dots, m_i$$

โดยที่  $\bar{x}_i^{(i)}$  คือค่าเฉลี่ยของตัวแปรที่  $i$  คำนวณมาจากเฉพาะข้อมูลของคนที่ตอบคำถามข้อที่  $i$  สัญลักษณ์  $(i)$  กำกับเอาไว้เพื่อแสดงว่าสนใจเฉพาะผู้ที่ตอบคำถามข้อที่  $i$

obs (v) คือบรรดาคำตอบของผู้ตอบรายที่  $v$  สำหรับคำถามต่าง ๆ เฉพาะข้อที่ตอบ

$i \in obs(v)$  หมายถึงเราสนใจเฉพาะคำตอบของทุกคำถามที่ตอบ คำถามที่ไม่ตอบไม่นำมาพิจารณา

6. **วิธี ICS (item correlation substitution)** วิธี ICS ดำเนินการโดยคำนวณหาสหสัมพันธ์ระหว่างตัวแปรโดยวิธี pairwise deletion จากค่าสหสัมพันธ์ให้ตรวจดูว่าตัวแปรที่เรากำลังพิจารณา มีความสัมพันธ์กับตัวแปรใดที่สุด เมื่อพบแล้วให้นำคำตอบของหน่วยสำรวจที่ตอบคำถามนั้นแทนที่ให้แก่ตัวแปรที่สัมพันธ์กันแต่มีข้อมูลสูญหาย ถ้าผู้ตอบไม่ตอบทั้งสองคำถาม ให้เปลี่ยนไปใช้คำตอบของตัวแปรที่สัมพันธ์กับตัวแปรที่กำลังพิจารณาในระดับรองลงไป

การพิจารณาตัวแปรตัวหนึ่งตัวใดหรือคู่หนึ่งคู่ใดเป็นเรื่องของการศึกษาความเห็นในระดับกลุ่ม ถ้าตัวแปรคู่ใดมีความสัมพันธ์กันสูงก็แสดงว่ากลุ่มมีแนวความเห็นต่อเรื่องทั้งสองนั้นไปในทางเดียวกัน น่าจะนำเอาข้อมูลจากตัวแปรที่คู่กันนั้นมาทดแทนกันได้

7 **วิธี HDD (hot-deck deterministic)** วิธี HDD ดำเนินการโดย คำนวณหาระยะทางระหว่างหน่วยสำรวจที่มีข้อมูลสูญหายกับหน่วยที่มีข้อมูลครบทุกตัวแปรตามแนวทางของ pairwise deletion แล้วนำข้อมูลจากหน่วยสำรวจที่สมบูรณ์ (คือตอบทุกคำถาม) ที่สัมพันธ์กับหน่วยที่มีข้อมูลสูญหายมากที่สุด (ระยะทางสั้นที่สุด) เป็นข้อมูลทดแทนถ้าระยะทางระหว่างหน่วยสำรวจที่มีข้อมูลสูญหายกับหน่วยที่มีข้อมูล

สมบุรณ์หลายหน่วยมีค่าเท่ากันให้เลือกใช้ข้อมูลจากหน่วยที่อยู่ใกล้กว่าเป็นข้อมูลทดแทน โดยมีความเชื่อว่าหน่วยที่มีเลขที่ใกล้เคียงกันคือหน่วยสำรวจในย่านหรือในพื้นที่ทางภูมิศาสตร์หรือกลุ่มประชากรย่อยเดียวกันน่าจะมีธรรมชาติคล้ายกันมากกว่าหน่วยที่ไกลออกไป

**8 วิธี HDR (hot-deck random)** วิธี HDR ดำเนินการเช่นเดียวกับวิธี HDD แตกต่างกันเพียงเล็กน้อยคือเมื่อพบวาระยะทางระหว่างหน่วยสำรวจที่มีข้อมูลสูญหายกับหน่วยที่มีข้อมูลสมบุรณ์หลายหน่วยมีค่าเท่ากันให้เลือกใช้หน่วยที่สมบุรณ์โดยสุ่มและใช้ข้อมูลจากหน่วยสมบุรณ์นั้นเป็นข้อมูลทดแทนให้แก่รายการที่สูญหายของหน่วยที่ไม่สมบุรณ์ โดยเชื่อว่าหน่วยสำรวจที่เราเข้าทำการสอบถามนั้นอยู่ในพื้นที่เดียวกัน หรือมีธรรมชาติคล้ายกัน หรือมีสภาพแวดล้อมอื่นคล้ายกัน ย่อมมีความเป็นไปได้ที่จะมีข้อมูลคล้ายกัน การเลือกหน่วยสมบุรณ์มาใช้เป็นหน่วยให้ข้อมูลจึงอาจเลือกมาจากหน่วยใดก็ได้

**9 วิธี HDD-IMS (hot-deck deterministic-item mean substitution)** วิธี HDD-IMS เป็นวิธีผสมคือใช้วิธีแทนที่ข้อมูลที่สูญหายด้วยค่าเฉลี่ยของตัวแปรเพื่อให้มีค่าสังเกตที่สมบุรณ์ครบทุกหน่วยเสียก่อน จากนั้นจึงคำนวณหาระยะทางระหว่างหน่วยสำรวจ แล้วตัดสินใจเลือกหน่วยให้ข้อมูลตามวิธี HDD วิธีนี้มีข้อดีที่ทุกหน่วยสำรวจมีสารสนเทศครบถ้วนตั้งแต่แรกก่อนที่จะคำนวณระยะทาง มีผลให้ค่าระยะทางคำนวณมาจากสารสนเทศที่มากกว่าวิธี HDD

**10 วิธี HDD-PMS (hot-deck deterministic-person mean substitution)** วิธี HDD-PMS เป็นวิธีผสมคือเป็นวิธีที่เราแทนที่ข้อมูลที่สูญหายด้วยค่าเฉลี่ยของหน่วยสำรวจนั่นเองเพื่อให้มีค่าสังเกตที่สมบุรณ์ครบทุกหน่วยก่อนที่จะดำเนินการเพื่อคำนวณหาระยะทาง การแทนที่ข้อมูลกรณีนี้ก็เพื่อต้องการให้หน่วยสำรวจมีสารสนเทศมากที่สุดเท่าที่จะมากได้สำหรับใช้คำนวณความสัมพันธ์ระหว่างหน่วยสำรวจต่อไป วิธีนี้แตกต่างจากวิธี HDD-IMS เฉพาะขั้นการแทนที่ข้อมูลเท่านั้น

**11 วิธี HDR-IMS (hot-deck random-item mean substitution)** วิธี HDR-IMS เป็นวิธีผสมคือเป็นวิธีที่เราแทนที่รายการข้อมูลที่สูญหายของหน่วยสำรวจใด ๆ ด้วยค่าเฉลี่ยของเฉพาะตัวแปรนั้น จากนั้นจึงคำนวณหาระยะทางระหว่างหน่วยสำรวจ แล้วจึงเลือกหน่วยให้ข้อมูลแก่หน่วยรับข้อมูลที่กำหนด โดยเลือกหน่วยให้ข้อมูลจากหน่วยที่มีระยะทางห่างจากหน่วยรับข้อมูลเป็นระยะทางที่สั้นที่สุด ถ้ามีระยะทางที่สั้นที่สุดเท่ากันหลายค่าให้เลือกขึ้นมาหนึ่งหน่วยโดยสุ่ม

**12 วิธี HDR-PMS (hot-deck random-person mean substitution)** วิธี HDR-PMS เป็นวิธีผสมคือเป็นวิธีที่เราแทนที่รายการข้อมูลที่มีค่าสูญหายของหน่วยสำรวจใด ๆ ด้วยค่าเฉลี่ยของข้อมูลจากหน่วยสำรวจนั้น แล้วค่อยคำนวณหาระยะทางระหว่างหน่วยสำรวจ จากนั้นจึงคัดเลือกหน่วยให้ข้อมูลจากระยะทางที่ได้ ถ้าหน่วยรับข้อมูลใดมีระยะทางห่างจากหน่วยให้ข้อมูลหลายหน่วยด้วยระยะทางสั้นที่สุดเท่า ๆ กัน ให้เลือกหน่วยให้ข้อมูลมาโดยสุ่มจากหน่วยเหล่านั้น

**13. วิธี HDD-Z (standardized HDD)** วิธี HDD-Z คือวิธี HDD ที่กระทำกับข้อมูลของทุกตัวแปรที่ได้แปลงเป็นคะแนนมาตรฐาน (Z) เอาไว้แล้ว การแปลงข้อมูลเป็นคะแนนมาตรฐานมีวัตถุประสงค์เพื่อปรับค่าตัวแปรที่วัดค่าด้วยมาตราที่แตกต่างกันให้เป็นข้อมูลในมาตรฐานเดียวกันคือ  $Z = \frac{X - \bar{X}}{SD}$  และเมื่อได้ดำเนินการตามวิธี HDD แล้วให้แปลงข้อมูลกลับสู่มาตราเดิมคือ  $Z = \bar{X} + SD * Z$

14. **วิธี HDR-Z (standardized HDR)** วิธี HDR-Z คือวิธี HDR ที่กระทำกับข้อมูลของทุกตัวแปรที่ได้แปลงค่าเป็นคะแนนมาตรฐานเอาไว้ก่อนแล้ว และเมื่อได้ดำเนินการตามวิธี HDR แล้วให้แปลงข้อมูลกลับสู่มาตราเดิม

15. **วิธี HDD-IMS-Z (standardized HDD-IMS)** วิธี HDD-IMS-Z คือวิธีผสมที่คัดเลือกหน่วยให้ข้อมูลและทดแทนข้อมูลที่สูญหายตามวิธี HDD-IMS โดยที่ก่อนดำเนินการจะต้องแปลงข้อมูลทุกรายการให้เป็นคะแนนมาตรฐานเสียก่อน และเมื่อได้ดำเนินการทดแทนข้อมูลเสร็จแล้วให้แปลงข้อมูลกลับคืนสู่มาตราเดิม

16. **วิธี HDD-PMS-Z (standardized HDD-PMS)** วิธี HDD-PMS-Z คือวิธีผสมที่คัดเลือกหน่วยให้ข้อมูลและทดแทนข้อมูลที่สูญหายตามวิธี HDD-PMS โดยที่ก่อนดำเนินการจะต้องแปลงข้อมูลทุกรายการให้เป็นคะแนนมาตรฐานเสียก่อน และเมื่อได้ดำเนินการทดแทนข้อมูลเสร็จแล้วให้แปลงข้อมูลกลับคืนสู่มาตราเดิม

17. **วิธี HDR-IMS-Z (standardized HDR-IMS)** วิธี HDR-IMS-Z คือวิธีผสมที่คัดเลือกหน่วยให้ข้อมูลและทดแทนข้อมูลที่สูญหายตามวิธี HDR-IMS โดยที่ก่อนดำเนินการจะต้องแปลงข้อมูลทุกรายการให้เป็นคะแนนมาตรฐานเสียก่อน และเมื่อได้ดำเนินการทดแทนข้อมูลเสร็จแล้วให้แปลงข้อมูลกลับคืนสู่มาตราเดิม

18. **วิธี HDR-PMS-Z (standardized HDR-PMS)** วิธี HDR-PMS-Z คือวิธีผสมที่คัดเลือกหน่วยให้ข้อมูลและทดแทนข้อมูลที่สูญหายตามวิธี HDD-IMS โดยที่ก่อนดำเนินการจะต้องแปลงข้อมูลทุกรายการให้เป็นคะแนนมาตรฐานเสียก่อน และเมื่อได้ดำเนินการทดแทนข้อมูลเสร็จแล้วให้แปลงข้อมูลกลับคืนสู่มาตราเดิม

## วิธีการวิเคราะห์

1. **วิธีทดลอง** การวิจัยครั้งนี้แยกการทดลองและการวิเคราะห์เป็น 2 แนวทางดังนี้

ก. **การทดลองวิเคราะห์เชิงประจักษ์ (empirical experiment)** คือการทดลองกับข้อมูลที่มีอยู่จริง ใช้สารสนเทศและข้อมูลที่มีอยู่เป็นเครื่องมือ

1. จากการศึกษาพบว่าเพศ อายุ การศึกษา รายได้ และอาชีพมีผลต่ออัตราการตอบแบบสอบถามซึ่งส่งผลต่อการสูญหายของข้อมูลโดยพบว่าหญิง ผู้มีการศึกษาน้อย ผู้มีสุขภาพไม่ดี และผู้สูงอายุจะมีอัตราการไม่ตอบสูง (DeLeeuw, 2001; Huisman, & van de Zouwen, 1998; Huisman, et al., 1998) โดยหญิงที่อยู่บ้านคนเดียวจะมีอัตราการไม่ตอบสูงมาก (Rucker, 1990 อ้างถึงใน Couper & Groves, 1996) แต่ก็พบว่าอายุอาจมีผลต่อการตอบหรือไม่ก็ได้ คือผู้สูงอายุอาจยินดีหรือไม่ยินดีร่วมมือขึ้นอยู่กับเรื่องที่สอบถามและสุขภาพผู้ตอบ (Couper & Groves, 1998) ขณะเดียวกันก็กลับพบว่าหญิงให้ความร่วมมือมากกว่า (Couper & Groves, 1998) และพบว่าผู้มีรายได้ต่ำ อาชีพไม่มั่นคง จะมีอัตราการไม่ตอบสูง (Demaio, 1980 อ้างถึงใน Rylander et al., 1995)

ด้วยเหตุนี้การวิจัยครั้งนี้จึงกำหนดให้ตัวแปรอิสระคือ เพศ อายุ การศึกษา รายได้ และอาชีพทั้งในกรณีการทดลองโดยใช้ข้อมูลจริงและกรณีข้อมูลเลียนแบบ สำหรับกรณีข้อมูลเลียนแบบจะกำหนดให้ตัวแปร

ตามเป็นตัวแปรในมาตราวัดแบบอันดับ (ordinal scale) ส่วนกรณีข้อมูลจริงจะกำหนดให้ตัวแปรตามเป็นตัวแปรในมาตราวัดแบบนามบัญญัติ (nominal scale)

2. กำหนดขนาดตัวอย่างเป็น 3 ระดับคือ 100, 200, 500 เหตุที่ไม่ใช้ขนาดตัวอย่างที่เล็กกว่านี้ เพราะขนาดตัวอย่างที่ใหญ่ขึ้นทำให้การทดแทนข้อมูลตามวิธี hot deck มีหน่วยผู้ให้ข้อมูล (donor) มากขึ้น ผลการศึกษาก็จะแม่นยำมากขึ้น (Roth & Switzer, 1995)

3. จากเพิ่มข้อมูลที่หน่วยวิเคราะห์ทุกหน่วยเป็นหน่วยสมบูรณ์ (complete case) กำหนดให้มีหน่วยสำรวจที่มีข้อมูลสูญหาย (missing case) ปะปนอยู่  $m_c$  % แต่ละหน่วยสำรวจดังกล่าวกำหนดให้มีตัวแปรที่ไม่ได้รับการตอบคำถาม (missing item) ปะปนอยู่  $m_i$  %

การตัดสินใจว่าหน่วยสำรวจใดเป็นหน่วยที่มีข้อมูลสูญหายกำหนดโดยใช้สมการถดถอยลอจิสติก

$$\Pr(\text{มีข้อมูลหาย}) = p = \frac{1}{1 + e^{-z}}$$

$$\text{โดยที่ } Z = 2.5 + 1 * \text{SEX} - 1 * \text{AGE} - 1 * \text{EDU} - 1 * \text{INC} + 1 * \text{OCC}$$

สมการข้างต้นคือสมการที่มีผลให้หญิง ผู้สูงอายุ ผู้มีการศึกษาน้อย ผู้มีรายได้น้อย และผู้มีอาชีพไม่มั่นคง มีโอกาสไม่ร่วมมือ (nonresponse) สูงกว่ากลุ่มอื่น (Huisman, 1997 และผลจากการจำลองแบบ) หน่วยสำรวจใดให้ค่า  $p \geq 0.5$  ถือว่าหน่วยสำรวจนั้นเป็นหน่วยสำรวจที่มีข้อมูลสูญหาย จากนั้นจึงกำหนดให้ข้อมูลของตัวแปรจากหน่วยสำรวจดังกล่าวหายไปชั่วคราวโดยสุ่มรวมทั้งสิ้น  $m_i$  % ด้วยการแทนที่ด้วยเลข 0 หรืออักขระใดๆ เช่นเครื่องหมาย “-” โดยที่  $m_c = 5, 10, 15, 20$  และ  $m_i = 10, 20, 30$  ตามลำดับ

4. สร้างข้อมูลทดแทนให้แก่ตัวแปรที่มีข้อมูลสูญหายไปชั่วคราวด้วยวิธีทดแทนข้อมูลวิธีต่างๆ รวม 18 วิธี เพิ่มข้อมูลที่ใช้เป็นแฟ้มข้อมูลเรื่อง “การสำรวจเกี่ยวกับอนามัยและสวัสดิการ พ.ศ. 2544” สำรวจโดยสำนักงานสถิติแห่งชาติ มีข้อมูลทั้งสิ้น 118,285 ระเบียบ คัดเอามาเฉพาะระเบียบที่สมบูรณ์ที่สุดคล้อยกับแบบสอบถามที่เลือกมาทดลองรวม 15 ข้อถาม จำนวน 83,641 ระเบียบ

5. กำหนดให้ทำการทดลองซ้ำแก่แต่ละเงื่อนไขการทดลองเงื่อนไขละ 100 ซ้ำ สำหรับ  $m_c = 5\%, 10\%, 15\%, 20\%$ ,  $m_i = 10\%, 20\%, 30\%$  และ  $n = 100, 200, 500$  รวมเป็นตัวอย่างทั้งสิ้น 3,600 กลุ่ม แล้วนำค่าสถิติจากแต่ละตัวอย่างมาถัวเฉลี่ยกันเพื่อแสดงคุณภาพของวิธีทดแทนข้อมูลแต่ละวิธี

**ข. การทดลองมอนติคาร์โล** คือการทดลองที่กระทำกับข้อมูลเลียนแบบ โดยใช้เลขสุ่มเป็นเครื่องมือ เลขสุ่มถูกสร้างขึ้นเป็น 2 ระยะคือ เลขสุ่มแสดงค่าตัวแปรอิสระคือ เพศ อายุ ระดับการศึกษา รายได้ และอาชีพ โดยกำหนดให้มีค่าที่เกิดขึ้นตามค่าของความน่าจะเป็นที่ปรากฏในตารางแจกแจงความถี่ และเลขสุ่มแสดงค่าตัวแปรตามซึ่งวัดในมาตราวัดแบบลิเกออร์ทกำหนดให้เกิดขึ้นด้วยสมการถดถอย

1. การสร้างค่าตัวแปรอิสระ ตัวแปรอิสระแสดงลักษณะทางประชากรเป็นตัวแปรกรุปรวม 5 ตัว คือ เพศ อายุ การศึกษา รายได้ และ อาชีพ การจัดโครงสร้างของตัวแปรอิสระอาศัยผลการกำหนดรหัสและร้อยละจากการสำรวจสำมะโนประชากรและการเคหะแห่งประเทศไทย พ.ศ. 2543

2. การสร้างค่าตัวแปรตามกำหนดให้มีตัวแปรตามทีวัดค่าตามมาตรวัดแบบลิเกอ์ทคือ  $V_1, V_2, V_3, V_4, V_5, V_6, V_7, V_8, V_9, V_{10}$  แต่ละตัวแปร มี 5 รหัส

ตัวแปรตามเป็นตัวแปรที่ขึ้นอยู่กับตัวแปรอิสระ ในที่นี้กำหนดให้ตัวแปรตามเป็นตัวแปรที่วัดค่าตามมาตรวัดแบบลิเกอ์ท ตัวแปรตามในที่นี้จึงถูกกำหนดให้มีค่าที่ผันแปรตามตัวแปรอิสระโดยกำหนดให้แปรไปตามระดับโครงสร้างความสัมพันธ์ภายในระหว่างตัวแปรอิสระซึ่งนักวิจัยสามารถกำหนดระดับสหสัมพันธ์ได้เองตามความต้องการ

กระบวนการสร้างค่าตัวแปรตามปรากฏดังนี้

ขั้นตอนที่ 1 กำหนดให้มีตัวแปรอิสระ 5 ตัว คือ เพศ (1, 2) อายุ (1, 2, 3, 4, 5) การศึกษา (1, 2, 3) รายได้ (1, 2, 3, 4, 5) อาชีพ (1, 2, 3, 4, 5)

ขั้นตอนที่ 2 ตัวแปรอิสระตามขั้นตอนที่ 1 ถูกสร้างโดยอาศัยสมการต่อไปนี้ (McDonald, 1975; Gibbons, 1981; Wichern & Churchill, 1978)

$$X_{ij} = (1 - \alpha^2)^{\frac{1}{2}} Z_{ij} + \alpha Z_{i6} ; j = 1, 2, 3 ; i = 1, 2, \dots, n$$

$$\text{และ } X_{ij} = (1 - \alpha_*^2)^{\frac{1}{2}} Z_{ij} + \alpha_* Z_{i6} ; j = 4, 5 ; i = 1, 2, \dots, n$$

โดยที่  $Z_{ij}$  เป็นตัวแปรมาตรฐานปกติที่เป็นอิสระต่อกันสร้างขึ้นโดยอาศัยขั้นตอนวิธีของมาร์ชาเกเลียและเบรีย (Wichern & Churchill 1978; Law & Kelton, 2000) และกำหนดให้  $\alpha^2 = .99^2, .90^2, .70^2$  และ  $\alpha_*^2 = .99^2, .90^2, .30^2, .10^2$  คือระดับสหสัมพันธ์ระหว่างตัวแปรอิสระและถูกปรับค่าให้มีรหัส ดังนี้

$$X_1 = \text{เพศ มีรหัส } 1, 2$$

$$X_2 = \text{อายุ มีรหัส } 1, 2, 3, 4, 5$$

$$X_3 = \text{การศึกษา มีรหัส } 1, 2, 3$$

$$X_4 = \text{รายได้ มีรหัส } 1, 2, 3, 4, 5$$

$$X_5 = \text{อาชีพ มีรหัส } 1, 2, 3, 4, 5$$

แนวทางการปรับค่าตัวแปรอิสระให้มีรหัสตรงตามความต้องการกำหนดขึ้นโดยการคำนวณหาค่าพื้นที่ใต้

โค้ง  $N(0,1)$  คือ  $\int_{-\infty}^{X_{ij}} f(z) dz$  แล้วเปลี่ยนเป็นรหัสโดยอาศัยค่าความน่าจะเป็นจากตารางแจกแจงความถี่ตาม

ผลการสำรวจสำมะโนประชากรและเคหะแห่งประเทศไทย พ.ศ. 2543

จากค่าสหสัมพันธ์เท่ากับ  $\alpha^2$  ระหว่างตัวแปรอิสระ  $X_1, X_2$  และ  $X_3$  และค่าสหสัมพันธ์เท่ากับ  $\alpha_*^2$  ระหว่าง  $X_4$  กับ  $X_5$  และจากค่าสหสัมพันธ์ร่วมเท่ากับ  $\alpha\alpha_*$  ระหว่าง  $X_1, X_2, X_3$  กับ  $X_4$  และ  $X_5$  จะพบว่าเมทริกซ์สหสัมพันธ์คือ  $X'X$  ขนาด  $5 \times 5$  ค่าสหสัมพันธ์  $\alpha^2$  ให้มีค่าเท่ากับ  $.99^2, .90^2, .70^2$  และกำหนดให้  $\alpha_*^2$  มีค่าเท่ากับ  $.99^2, .90^2, .30^2$  และ  $.10^2$  ก็จะได้เมทริกซ์สหสัมพันธ์ที่มีค่าเป็นตัวเลขซึ่งสามารถคำนวณหาค่าไอเกน (eigen value) และดัชนีวัดระดับภาวะร่วมเส้นตรงพหุ (เรียกว่า spectral condition number,  $R_m = \lambda_{\max} / \lambda_{\min}$ ) ถ้าดัชนีวัดระดับภาวะร่วมเส้นตรงพหุมีค่าสูงแสดงว่ากลุ่มตัวแปรอิสระมีความเกี่ยวข้องกันสูง ถ้าดัชนีวัดระดับภาวะร่วมเส้นตรงพหุมีค่าต่ำแสดงว่ากลุ่มตัวแปรอิสระเกี่ยวข้องกันน้อย การศึกษาแยกเป็น 5 กรณีคือกรณีมีภาวะร่วมเส้นตรงพหุสูงมากจนถึงกรณีมีภาวะร่วมเส้นตรงพหุต่ำ

พบว่าในแต่ละกรณีมีค่าไอเกน 5 ค่า ค่าไอเกนแต่ละค่า นำสู่ไอเกนเวกเตอร์ (eigen vector) ขนาด  $5 \times 1$  หนึ่งเวกเตอร์เสมอ สมาชิกของไอเกนเวกเตอร์ปรับปกติ (normalized eigen vector) ที่สอดคล้องกับค่าไอเกนต่ำสุด ( $\lambda_{\min}$ ) ใช้เป็นค่าสัมประสิทธิ์  $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5$  (Gibbons, 1981; Wichern & Churchill, 1978; McDonald & Galarneau, 1975) ในที่นี้เลือกใช้เฉพาะไอเกนเวกเตอร์ปรับปกติที่สอดคล้องกับค่าไอเกน  $\lambda_{\min}$  เพราะมีความประสงค์จะศึกษากรณีที่เสี่ยงต่อปัญหาภาวะร่วมเส้นตรงพหุสูงที่สุด เพราะหากไม่มีความเสียหายใดเกิดขึ้นกับกรณีนี้ก็คงไม่เกิดความเสียหายใดกับกรณีที่มีความเสี่ยงต่อปัญหาภาวะร่วมเส้นตรงพหุต่ำกว่านี้

ตัวแปรตามคือทัศนคติต่อเรื่องใดเรื่องหนึ่งจึงมีค่าที่สามารถสร้างขึ้นได้ดังนี้คือ

$$Y_i = \beta_{0s} + \beta_{1s}X_{i1} + \beta_{2s}X_{i2} + \beta_{3s}X_{i3} + \beta_{4s}X_{i4} + \beta_{5s}X_{i5} + e_i; i = 1, 2, \dots, n$$

กำหนดให้  $\beta_{0s}$  มีค่าเท่ากับ 0 และกำหนดให้  $e_i$  แจกแจงแบบปกติมาตรฐาน  $N(0,1)$  (Gibbons, 1981)

ตัวแปรตามถูกปรับรหัสให้เป็นจำนวนนับคือ 1, 2, 3, 4, 5 ตามกฎความน่าจะเป็นเชิงประจักษ์ (empirical rule of probability) โดยกำหนดให้แต่ละรหัสเกิดขึ้นด้วยความน่าจะเป็นเท่ากัน และให้รหัส 1, 2, 3, 4, 5 แสดงระดับทัศนคติเป็นข้อความว่า ไม่เห็นด้วยอย่างยิ่ง ไม่เห็นด้วย ไม่มีความเห็น เห็นด้วย และเห็นด้วยอย่างยิ่ง ตามลำดับ

ขั้นตอนที่ 3. กำหนดให้ข้อมูลสูญหาย การกำหนดให้ข้อมูลสูญหายใช้รูปแบบเดียวกับกรณีการทดลองเชิงประจักษ์คือตัดสินว่าหน่วยใดเป็นหน่วยที่มีข้อมูลสูญหายกำหนดโดยใช้ค่าความน่าจะเป็นจากสมการถดถอยลอจิสติก ดังนี้

$$\Pr(\text{มีข้อมูลหาย}) = p = \frac{1}{1 + e^{-z}}$$

$$\text{โดยที่ } Z = 2.5 + 1 * \text{SEX} - 1 * \text{AGE} - 1 * \text{EDU} - 1 * \text{INC} + 1 * \text{OCC}$$

ซึ่งจะมีผลให้หญิง ผู้สูงอายุ ผู้มีการศึกษาน้อย ผู้มีรายได้น้อย และผู้มีอาชีพไม่มั่นคง มีโอกาสไม่ให้ความร่วมมือสูงกว่ากลุ่มอื่น หน่วยสำรวจใดที่ให้ค่า  $p \geq 0.5$  ถือว่าหน่วยนั้นเป็นหน่วยที่มีข้อมูลสูญหาย

กำหนดให้  $m_c = 5\%, 10\%, 15\%, 20\%$ ,  $m_i = 10\%, 20\%, 30\%$  ระดับความเกี่ยวข้องกันระหว่างตัวแปรอิสระ (multicollinearity level,  $R_m$ ) รวม 5 ระดับ ทุกกรณีศึกษาเฉพาะกรณีวิกฤติคือ  $\lambda_{\min}$  เท่านั้น ขนาดตัวอย่าง  $n = 100, 200, 500$  เพื่อแสดงสถานการณ์ของตัวอย่างขนาดเล็ก ขนาดปานกลาง และขนาดใหญ่ กำหนดให้มีจำนวนการทดลองซ้ำเท่ากับ 100 ซ้ำทุกเงื่อนไขของ  $m_c \times m_i \times R_m \times n$  รวมเป็นตัวอย่างทั้งสิ้น 18,000 กลุ่ม แล้วนำค่าสถิติมาเฉลี่ยกันเป็นค่าสถิติที่ใช้แสดงผลการทดลองของแต่ละเงื่อนไข

## 2. สถิติตัดสินใจ

การตัดสินใจว่าวิธีทดแทนข้อมูล (MDT) แบบใดเหมาะสมที่สุดให้พิจารณาจากสถิติ 4 ตัวต่อไปนี้ (Roth & Stwitzer, 1995; Huisman, 1997)

$$1) \text{MSE} = \frac{1}{n} \sum (\text{residual})^2 \text{ โดยที่ residual} = \text{ค่าจริง} - \text{ค่าทดแทน}; \text{MSE} \geq 0$$

$$2) \text{Bias} = \frac{1}{n} \sqrt{\sum (\text{residual})} \text{ โดยที่ residual} = \text{ค่าจริง} - \text{ค่าทดแทน}; \text{Bias} \geq 0$$

$$3) \alpha = \frac{k}{k-1} \frac{s^2 - \sum_i^k s_i^2}{s^2} \text{ โดยที่ } k = \text{จำนวนข้อถาม}; 0 \leq \alpha \leq 1 \text{ วิธีทดแทนข้อมูลใดให้ค่าเฉลี่ย}$$

ของ  $\alpha$  สูงกว่าแสดงว่าเป็นวิธีทดแทนข้อมูลที่สอดคล้อง (consistent) กับค่าจริงมากกว่า

$$4) r_{\text{old,new}} = \frac{\sum xy}{\sqrt{\sum x^2 \sum y^2}}; 1 \leq r_{\text{old,new}} \leq 1$$

X คือข้อมูลชุดเดิมและ Y คือข้อมูลชุดใหม่ที่ทดแทนค่าสูญหายด้วยวิธีทดแทนแบบใดแบบหนึ่ง ถ้าวิธีทดแทนข้อมูลใดให้ค่า r สูงและเป็นปริมาณที่เป็นบวกแสดงว่าวิธีทดแทนข้อมูลวิธีนั้นให้ค่าประมาณที่สอดคล้องกับค่าจริงมากกว่า

## สรุปผล

1. สถิติตัดสินใจ MSE bias และ r ให้ผลลัพธ์สอดคล้องกันคือ ( $\alpha$  ไม่แสดงความแตกต่าง)

(1) ในกรณีของข้อมูลจริง วิธีทดแทนข้อมูลสูญหาย (MDT) ที่ดีที่สุดคือวิธี IMS และ HDD-IMS และวิธี HDD-IMS-Z

(2) ในกรณีของข้อมูลเลียนแบบ วิธีทดแทนข้อมูลสูญหาย (MDT) ที่ดีที่สุดคือวิธี PMS และ HDD-PMS และวิธี HDD-PMS-Z

เหตุที่ผลการศึกษาไม่ตรงกันอาจเป็นเพราะการทดลองทั้ง 2 กรณีใช้ข้อมูลต่างมาตรากัน โดยเฉพาะอย่างยิ่งคือความแตกต่างกันในจำนวนตัวเลือก การทดลองมอนต์คาร์โล ทดลองกับข้อมูลในมาตราอันดับ 5 จุด (5 point LIKERT scale) ขณะที่การวิเคราะห์เชิงประจักษ์ทดลองกับข้อมูลในมาตราแบบบัญญัติ (nominal scale) ที่มีตัวเลือก 2 ถึง 11 ตัวเลือก ความแตกต่างระหว่างมาตราอาจไม่มีผลกระทบเพราะไม่ว่าจะใช้มาตราใดค่าของตัวแปรก็คือรหัสสมมุติซึ่งเป็นตัวเลขที่นักวิจัยกำหนดขึ้นเอง และสามารถกำหนดให้มีค่าตรงกันหรือแตกต่างกันได้ แต่จำนวนค่าที่เป็นไปได้ที่ไม่เท่ากันอาจมีผลกระทบ

2. ผลการศึกษาในข้อ 1. ไม่ผูกพันกับขนาดตัวอย่าง (n) อัตราข้อมูลสูญหาย (missing case,  $m_c$  และ missing item,  $m_i$ ) และระดับภาวะร่วมเส้นตรงพหุ (multicollinearity level) หมายความว่าไม่ว่าขนาดตัวอย่างจะเป็นขนาดใด ไม่ว่าจะมึระดับของภาวะร่วมเส้นตรงพหุสูงหรือต่ำอย่างไร และไม่ว่าจะมีอัตราการสูญหายของข้อมูลเท่าไร เครื่องมือทดแทนข้อมูลที่ดีที่จะสามารถใช้ได้ทั่วไปคือวิธี IMS วิธี HDD-IMS วิธี HDD-IMS-Z สำหรับข้อมูลที่วัดค่าตามมาตราแบบบัญญัติ วิธี PMS วิธี HDD-PMS และวิธี HDD-PMS-Z สำหรับข้อมูลที่วัดค่าในมาตราอันดับแบบลิกเกอร์ท์

3. ผลการศึกษาคั้งนี้สอดคล้องกับผลการศึกษาของฮิวส์แมน (Huisman, 1997) ที่พบว่าวิธี IMS และวิธี PMS เป็นวิธีที่ดีและวิธี PMS ดีกว่าวิธี IMS และควรมีข้อมูลสูญหายไม่เกิน 20% รวมทั้งควรมีตัวอย่างขนาดใหญ่ (Downey & King, 1998) ขณะที่ผลการศึกษาคั้งนี้พบว่าวิธี IMS เหมาะสมกว่าวิธี

PMS เมื่อใช้กับข้อมูลในมาตรฐานบัญชีแต่วิธี PMS เหมาะสมกว่าวิธี IMS เมื่อใช้กับข้อมูลในมาตรฐานฉบับแบบลิเกอาร์ท แต่ไม่ผูกพันกับอัตราการสูญหายของข้อมูลและขนาดตัวอย่าง

4.วิธี IMS ผสมกับวิธี HDD หรือ HDD-Z กับวิธี PMS ผสมกับวิธี HDD หรือ HDD-Z เป็นวิธีทดแทนข้อมูลที่ควรเลือกใช้เพราะมีคุณภาพดีในระดับที่น่าพึงพอใจ การศึกษาครั้งนี้สอดคล้องกับผลการศึกษาของฮิวส์แมน (Huisman, 1997) ไกเซอร์ (Kaiser, 2003) และ สไตรค์ (Strike, 2001) ที่พบว่าวิธี HDD และ HDD-Z เป็นวิธีทดแทนข้อมูลสูญหายที่ดีที่สุดและวิธี HDD-Z เป็นวิธีที่แม่นยำที่สุด

การพิจารณาเลือกตัวแบบทดแทนจึงขึ้นอยู่กับมาตรวัดค่าตัวแปร ความง่ายในการใช้งานและตัวนักวิจัยเอง ถ้าผู้วิจัยต้องการความสะดวก ใช้ง่าย เข้าใจง่าย ให้เลือกใช้วิธี IMS กรณีเมื่อข้อมูลนั้นวัดค่าตามมาตรฐานบัญชี และให้เลือกใช้วิธี PMS เมื่อข้อมูลนั้นวัดค่าตามมาตรฐานฉบับแบบลิเกอาร์ท ถ้าต้องการความแม่นยำให้ใช้วิธีผสมคือวิธีทดแทนข้อมูลแบบ IMS หรือ PMS ผสมกับวิธี HDD หรือ HDD-Z แต่โดยข้อเท็จจริงแล้ววิธีทดแทนข้อมูลในกลุ่มเดียวกันมีคุณภาพไม่แตกต่างกันมากนัก

- Churchill, G.A. and Wichern, D.W. (1978). A Comparison of Ridge Estimators, *Techometrics*, 20:301 - 311
- Cohn, M. S. (1982). Factor Analysis of Incomplete Data: Comparison of Listwise Deletion, Pairwise Deletion  
Missing Data Imputation in Social Science Research: A Simple Model Simulation รองศาสตราจารย์ ดร. มนต์รี พิริยะกุล  
And Insertion of Missing Data Estimates Based on Principal Component, Ph.D.--Temple<sup>15</sup>  
University, *Dissertation Abstract International*.
- Couper, Mick P. and Stinson Linda L. (1999). Completion of Self Administered Questionnaires in a  
Sex Survey, *Jour. of Sex Research*.
- Couper, Mick P. and Robert, M. Groves (1996). Household-Level Determinant of Survey Nonresponse. *New  
Directions for Evaluation*.
- Couper, Mick P. and Robert M. Groves (1998). **Nonresponse in Household Interview Survey**, N.Y.: John  
Wiley & Sons.
- DeLeeuw, E.D. (2001). Reducing Missing Data in Survey: An Overview of Methods, *Quality & Quantity*.
- DeSilvio, Michelle Lee (1999). A Variance Ratio Statistics for Assessing the Missing Data Mechanism: An  
Empirical Study, Ph. D. -Tulane University, *Dissertation Abstracts International*.
- Downey R.G. and King, C. (1998). Missing Data in Likert Ratings: A Comparison of Replacement  
Methods, *The Journal of General Psychology*.
- Gibbons, D.G. (1981). A Simulation Study of Some Ridge Estimators, *Journal of the American  
Statistical Association*, 76: 131-139.
- Heeringa, Steven George (2000). Multivariate Imputation of Coarsened Survey on Household Wealth, Ph.D.-  
University of Michigan. *Dissertation Abstract International*.
- Huisman, M. (1997). Imputation of Missing Item Responses: Some Simple Techniques, In Huisman, M.  
(edited). *Nonresponse: Occurrence causes and Imputation of Missing Answers to Test Item*. DSWO  
Press, Lieden University, The Netherlands.
- Huisman, M. (1998). Missing Data in Behavioral Science, *Kwantitative Methoden*, In Huisman, M.  
(edited). *Item Nonrespon : Occurrence causes, and Imputation of Missing Answers to  
Test Item*. DSWO Press, Lieden University, The Netherlands.
- Huisman, M., Krol, B. and Van Sonderen, F.L.P. (1998). Handling Missing Data by Re-approaching  
Nonrespondent, *Quality & Quantity*. In Huisman, M. (edited). *Item Nonresponse: Occurrence  
causes, and Imputation of Missing Answers to Test Item*, DSWO Press, Lieden University,  
The Netherlands.
- Ibrahim, J.G. (1988). Incomplete Data In Generalized Linear Model, Ph.D. - University of Minnesota,  
*Dissertation Abstracts International*.
- Jeon, Yoon Sook (1998). Inference in Structural Models With Missing Data, Ph.D. - Iowa State  
University. *Dissertation Abstracts International*.
- Kim Jae Kwang (2000). Variance Estimation after Imputation, Ph.D. - Iowa State University. *Dissertation  
Abstract International*.
- King, C.V. and Downey, R.G. (1998). Missing Data in Likert Ratings: A Comparison of Replacement  
Methods, *The Journal of General Psychology*, 125: 175-191.
- Krohn, Karl Ruben (1992). Incomplete Data in Surveys of Human Populations: A Review of Sources and Solution  
(volume I and II), Ph.D.-University of Minnesota, *Dissertation Abstract International*.
- Little, R. J. A. and Rubin, D. (1987). **Statistical Analysis with Missing Data**, N.Y.: John Wiley & Sons.

## ประวัตินักวิจัย

ชื่อ รองศาสตราจารย์มนตรี พิริยะกุล  
วัน เดือน ปี เกิด 13 ธันวาคม 2489  
วุฒิการศึกษา กศ.บ (คณิตศาสตร์) มหาวิทยาลัยนเรศวร 2512 พบ.ม (สถิติประยุกต์) สถาบัน  
บัณฑิตพัฒนบริหารศาสตร์ 2515

สถานที่ทำงานปัจจุบัน ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยรามคำแหง  
ผลงานทางวิชาการ

### ตำรา

- (1) มนตรี พิริยะกุล. *คณิตศาสตร์สำหรับนักสถิติ ตอนที่ 1 : เมตริกซ์และพีชคณิตเชิงเส้น*. กรุงเทพมหานคร: มหาวิทยาลัยรามคำแหง, 2536
- (2) มนตรี พิริยะกุล. *เทคนิคการสำรวจด้วยกลุ่มตัวอย่าง*. กรุงเทพมหานคร: มหาวิทยาลัยรามคำแหง, 2543.
- (3) มนตรี พิริยะกุล. *ทฤษฎีสถิติ 2*. กรุงเทพมหานคร: มหาวิทยาลัยรามคำแหง, 2546
- (4) มนตรี พิริยะกุล. *วิลิแคลค์*. กรุงเทพมหานคร: สำนักพิมพ์ยูไนเต็ดบุคส์, 2528.
- (5) มนตรี พิริยะกุล. *เทคนิคการวิเคราะห์สมการถดถอย*. กรุงเทพมหานคร: มหาวิทยาลัยรามคำแหง, 2544.
- (6) มนตรี พิริยะกุล. *การวิเคราะห์ทางสถิติของตัวแปรพหุ 2*. กรุงเทพมหานคร: มหาวิทยาลัยรามคำแหง, 2545.
- (7) มนตรี พิริยะกุล และคณะ. *สถิติเบื้องต้น*. กรุงเทพมหานคร: มหาวิทยาลัยรามคำแหง, 2530
- (8) มนตรี พิริยะกุล และ ระพีพรรณ พิริยะกุล. *ภาษาซี*. กรุงเทพมหานคร: อรุณการพิมพ์, 2532.
- (9) มนตรี พิริยะกุล. *เทคนิคการพยากรณ์ขั้นสูง*, กรุงเทพมหานคร: มหาวิทยาลัยรามคำแหง, 2546.
- (10) เฉลียว บุรีภักดี, มนตรี พิริยะกุล และ ชุศรี อิศราภรณ์. *ประชากรศึกษา*. กรุงเทพมหานคร: เจริญวิทย์การพิมพ์, 2520.

### ผลงานวิจัย

- (1) มนตรี พิริยะกุล. *การวิเคราะห์ปัญหามัลติโคลิเนียริตีในสมการถดถอยพหุคูณเมื่อมีค่าสังเกตสูญหาย*. รายงานการวิจัยเสนอต่อสำนักงานคณะกรรมการการวิจัยแห่งชาติและมหาวิทยาลัยรามคำแหง, 2528
- (2) มนตรี พิริยะกุล. *โครงสร้างการจัดสรรค่าใช้จ่ายของครัวเรือนภายใต้ภาวะความเปลี่ยนแปลงราคาน้ำมัน*. รายงานการวิจัยเสนอต่อสำนักงานคณะกรรมการการวิจัยแห่งชาติ, 2528.
- (3) มนตรี พิริยะกุล และ ระพีพรรณ พิริยะกุล. *การประเมินค่าสังเกตที่สูญหายในการวิเคราะห์ความถดถอยอย่างง่าย*. รายงานการวิจัยเสนอต่อสำนักงานคณะกรรมการการวิจัยแห่งชาติ, 2524.
- (4) ระพีพรรณ พิริยะกุล และ มนตรี พิริยะกุล. *การตรวจสอบและแก้ไขข้อผิดพลาดในการประมวลผลข้อมูล*. รายงานการวิจัยเสนอต่อสำนักงานคณะกรรมการการวิจัยแห่งชาติ, 2523.
- (5) มนตรี พิริยะกุล. *ตัวแบบการทดแทนข้อมูลในการวิจัยทางสังคมศาสตร์: การจำลองแบบกรณีตัวแบบอย่างง่าย*. รายงานการวิจัยเสนอต่อมหาวิทยาลัยรามคำแหง, 2548.

## ประสบการณ์

### 1. ประสบการณ์ทางวิชาการ

- (1) กรรมการผู้ทรงคุณวุฒิพิจารณาผลงานทางวิชาการ มหาวิทยาลัยศิลปากร นครปฐม
- (2) กรรมการผู้ทรงคุณวุฒิพิจารณาผลงานทางวิชาการ สถาบันเทคโนโลยีพระจอมเกล้า พระนครเหนือ
- (3) กรรมการผู้ทรงคุณวุฒิพิจารณาผลงานทางวิชาการ สถาบันเทคโนโลยีนานาชาติ
- (4) กรรมการที่ปรึกษาและประเมินคุณภาพตำรา ทบวงมหาวิทยาลัย
- (5) อาจารย์ผู้บรรยายวิชา MIS หลักสูตร MBA คณะบริหารธุรกิจ มหาวิทยาลัยรามคำแหง
- (6) อาจารย์ผู้บรรยายวิชา Simulation Technique วิชา Advanced Forecasting Technique วิชา Sampling Theory หลักสูตร M.S. (DSS) คณะวิทยาศาสตร์ มหาวิทยาลัยรามคำแหง
- (7) อาจารย์ผู้บรรยายวิชา Regression Analysis, Introduction to Simulation, Theory of Statistics, Theory of Probability, Sample Survey Techniques, Statistical Package Programs

### 2. ประสบการณ์ด้านการบริหาร

- (1) หัวหน้าภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยรามคำแหง 2530-2536
- (2) รองคณบดี คณะวิศวกรรมศาสตร์ มหาวิทยาลัยรามคำแหง 2540-2546
- (3) กรรมการอำนวยการคณะวิศวกรรมศาสตร์ มหาวิทยาลัยรามคำแหง
- (4) ประธานกรรมการวิชาการ คณะวิศวกรรมศาสตร์ 2540-2543
- (5) กรรมการวิชาการคณะวิทยาศาสตร์ 2527-2530

### 3. ประสบการณ์ด้านอื่น

- (1) ที่ปรึกษา บริษัท เซวาเลียร์ จำกัด (การสำรวจพฤติกรรมผู้บริโภคและการพยากรณ์ความต้องการใช้เครื่องเพชรเจียร, 2535)
- (2) ที่ปรึกษา บริษัท ขนส่ง จำกัด (การสำรวจความพึงพอใจของผู้โดยสาร, 2542)
- (3) ที่ปรึกษากระทรวงพลังงาน (การสำรวจการรับรู้ ความรู้ความเข้าใจ ความพึงพอใจ และการมีส่วนร่วมของประชาชนในมาตรการพลังงาน และการสำรวจความพึงพอใจในการปฏิบัติงานของข้าราชการ กระทรวงพลังงาน, 2547).

## 2. ตัวแปรที่ใช้ในกรณีวิเคราะห์เชิงประจักษ์

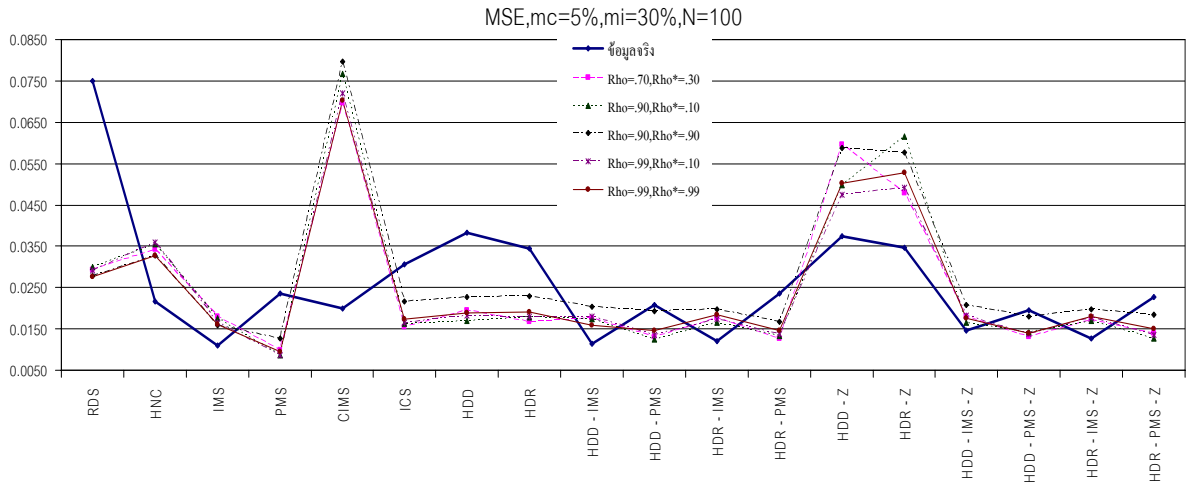
### ก. ตัวแปรอิสระประกอบด้วย

- (1). เพศ
- (2). อายุ
- (3). สถานภาพสมรส
- (4). การศึกษาสูงสุด
- (5). อาชีพ

### ข. ตัวแปรตามประกอบด้วย

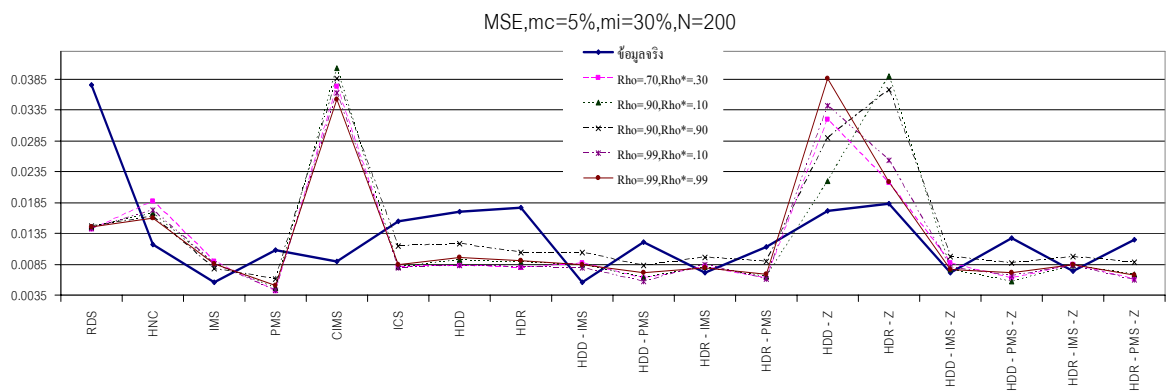
- (6) ระหว่าง 12 เดือนก่อนวันสัมภาษณ์ท่านเคยเจ็บป่วยจนต้องนอนรักษาตัวในโรงพยาบาลหรือไม่
- (7) สวัสดิการค่ารักษาพยาบาลที่.....(ชื่อ).....ได้รับเป็นสิทธิของใคร
- (8) ระหว่าง 12 เดือนก่อนวันสัมภาษณ์.....(ชื่อ).....เคยได้รับการตรวจร่างกายประจำปีหรือไม่
- (9) ระหว่าง 3 เดือนก่อนวันสัมภาษณ์.....(ชื่อ).....ออกกำลังกายหรือไม่
- (10) ถ้า.....(ชื่อ).....ต้องซื้ออาหารรับประทานจะคำนึงถึงสิ่งใดเป็นอันดับแรก
- (11) ปัจจุบัน.....(ชื่อ).....สูบบุหรี่หรือไม่
- (12) ปัจจุบัน.....(ชื่อ).....ดื่มสุราหรือเครื่องดื่มมีแอลกอฮอล์หรือไม่
- (13) ตามปกติ.....(ชื่อ).....ขับรถหรือไม่
- (14) ปกติ.....(ชื่อ).....ขณะขับหรือโดยสารรถยนต์ตอนหน้า ใช้เข็มขัดหรือไม่
- (15) ปกติ.....(ชื่อ).....ขณะขับหรือโดยสารรถจักรยานยนต์ได้สวมหมวกกันน็อคหรือไม่

### 3. ตัวอย่างผลการวิจัย



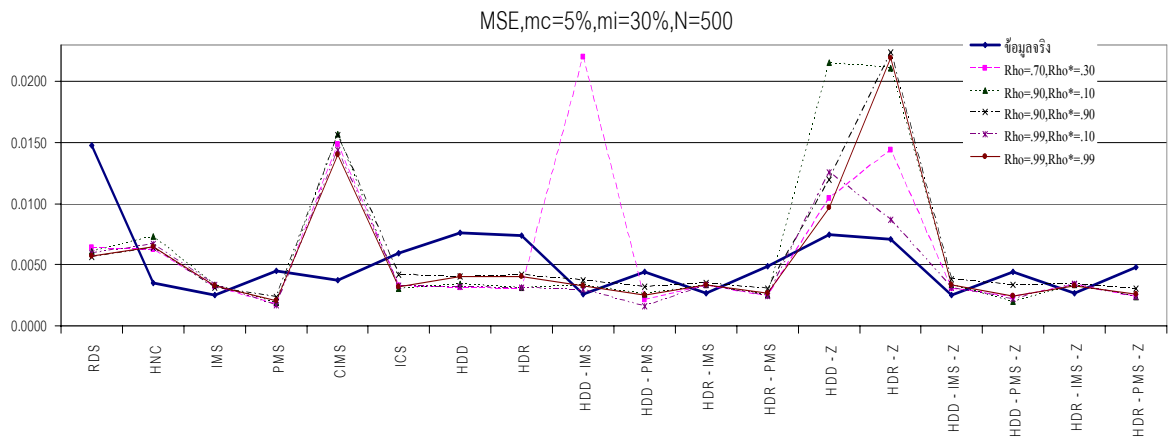
ภาพ 1

กราฟค่าเฉลี่ยของ MSE จากข้อมูลจริงและข้อมูลเลียนแบบ 5 กรณี เมื่อกำหนดให้  $m_c=5\%$ ,  $m_i=30\%$ ,  $N=100$



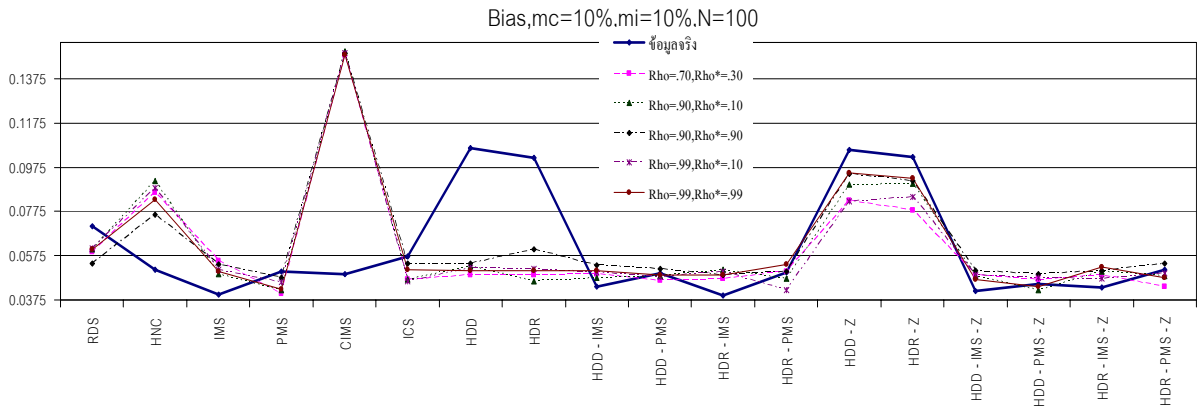
ภาพ 2

กราฟค่าเฉลี่ยของ MSE จากข้อมูลจริงและข้อมูลเลียนแบบ 5 กรณี เมื่อกำหนดให้  $m_c=5\%$ ,  $m_i=30\%$ ,  $N=200$



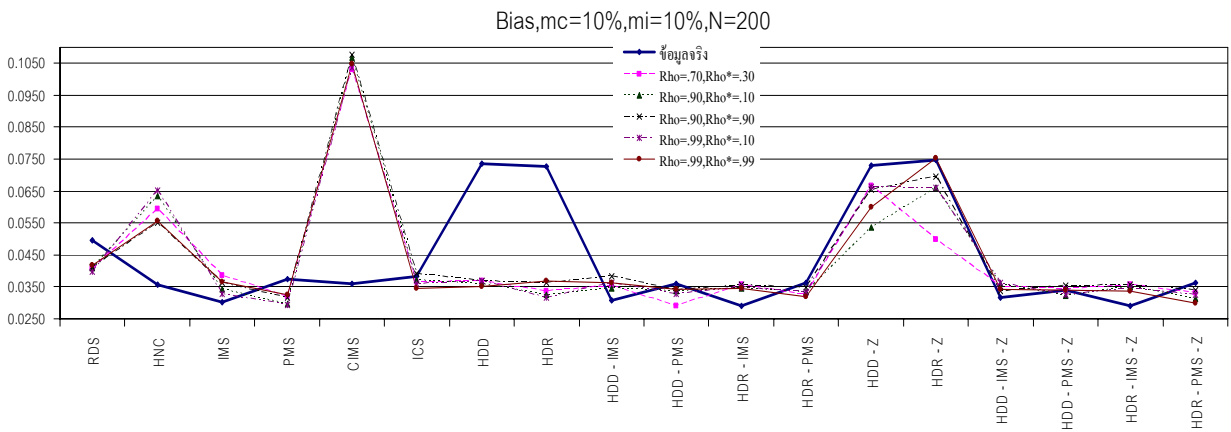
ภาพ 3

กราฟค่าเฉลี่ยของ MSE จากข้อมูลจริงและข้อมูลเลียนแบบ 5 กรณี เมื่อกำหนดให้  $m_c=5\%$ ,  $m_i=30\%$ ,  $N=500$



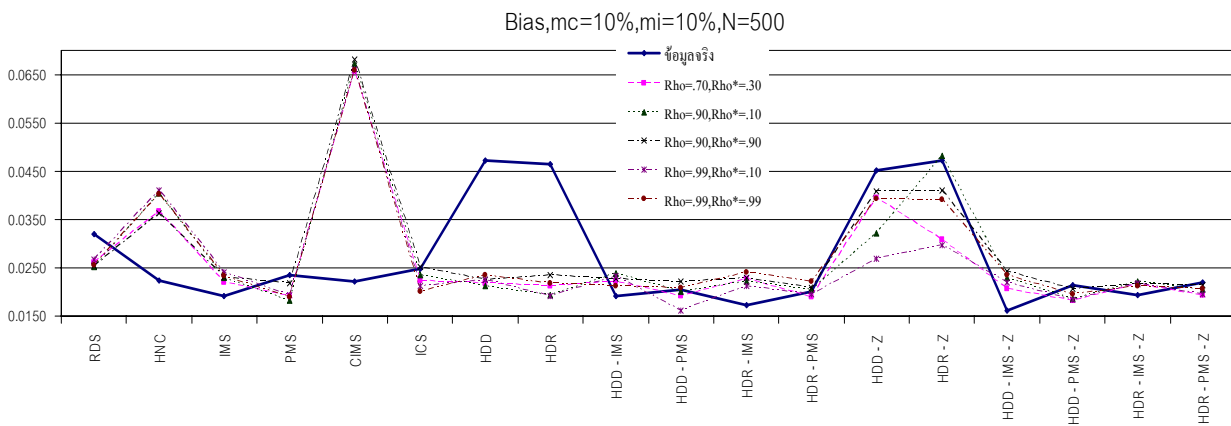
ภาพ 4

กราฟค่าเฉลี่ยของ bias จากข้อมูลจริงและข้อมูลเลียนแบบ 5 กรณี เมื่อกำหนดให้  $m_c=10\%$ ,  $m_i=10\%$ ,  $N=100$



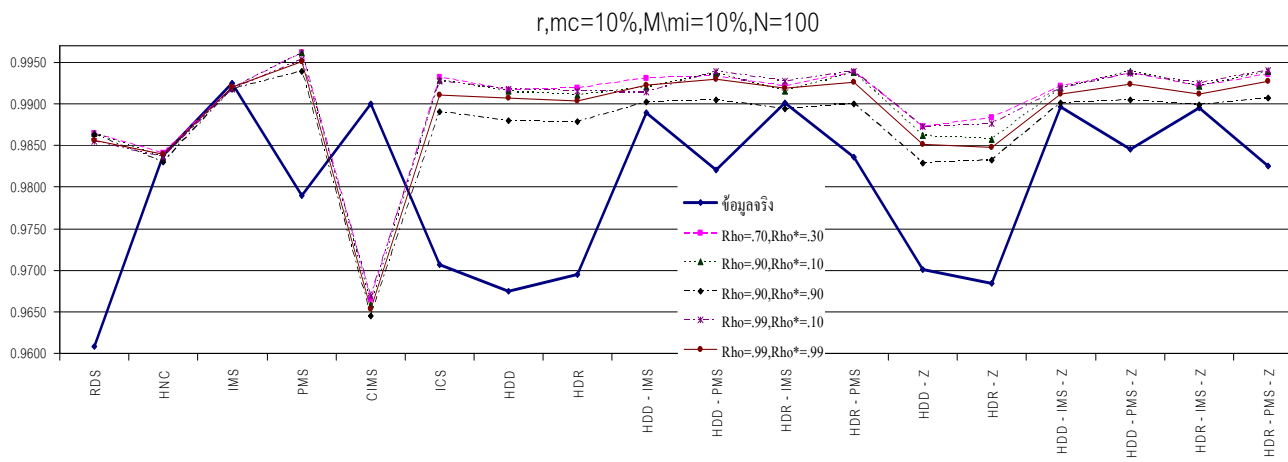
ภาพ 5

กราฟค่าเฉลี่ยของ bias จากข้อมูลจริงและข้อมูลเลียนแบบ 5 กรณี เมื่อกำหนดให้  $m_c=10\%$ ,  $m_i=10\%$ ,  $N=200$



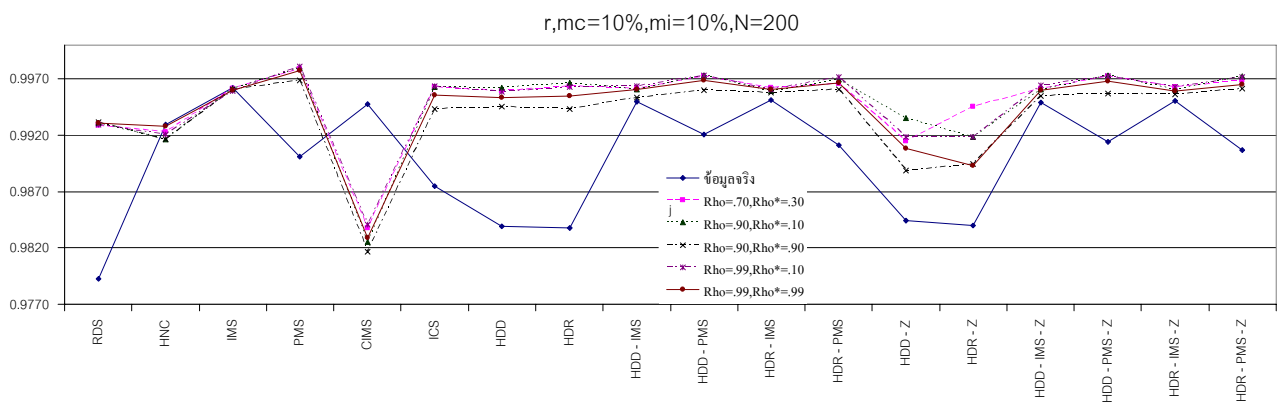
ภาพ 6

กราฟค่าเฉลี่ยของ MSE จากข้อมูลจริงและข้อมูลเลียนแบบ 5 กรณี เมื่อกำหนดให้  $m_c=10\%$ ,  $m_i=10\%$ ,  $N=500$



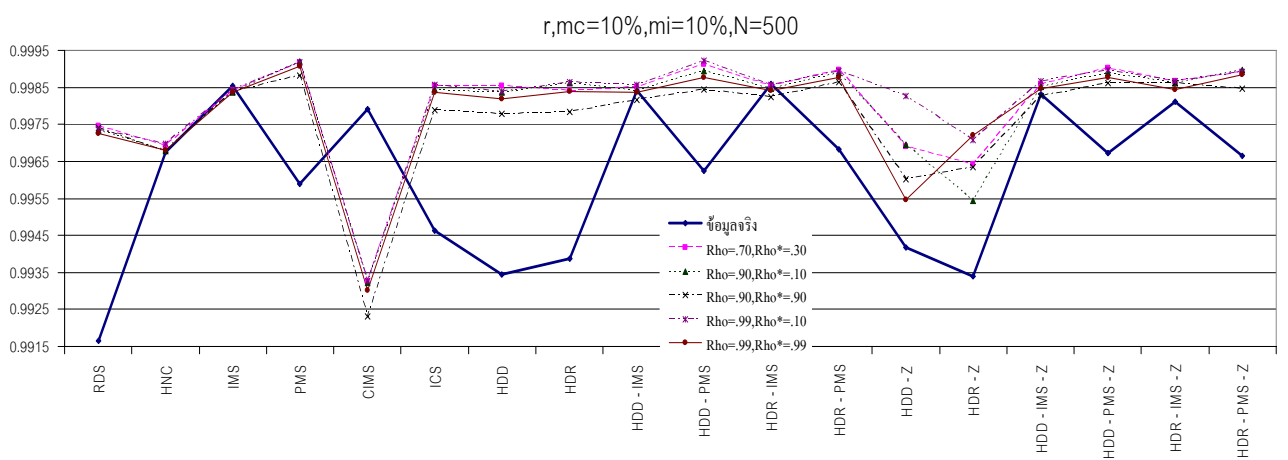
ภาพ 7

กราฟค่าเฉลี่ยของ r จากข้อมูลจริงและข้อมูลเลียนแบบ 5 กรณี เมื่อกำหนดให้  $m_c=10\%$ ,  $m_i=10\%$ ,  $N=100$



ภาพ 8

กราฟค่าเฉลี่ยของ r จากข้อมูลจริงและข้อมูลเลียนแบบ 5 กรณี เมื่อกำหนดให้  $m_c=10\%$ ,  $m_i=10\%$ ,  $N=200$



ภาพ 9

กราฟค่าเฉลี่ยของ r จากข้อมูลจริงและข้อมูลเลียนแบบ 5 กรณี เมื่อกำหนดให้  $m_c=10\%$ ,  $m_i=10\%$ ,  $N=500$