

น้ำทิพย์ พนมไทย 2554: การศึกษาเปรียบเทียบวิธีการประมาณค่าข้อมูลสูญหาย  
สำหรับข้อมูลตัวแปรพหุ ปรินญาวิทยาศาสตร์มหาบัณฑิต (สจิติ) สาขาสถิติ ภาควิชาสถิติ  
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก: ผู้ช่วยศาสตราจารย์บุญอ้อม โจนมทิ, Ph.D. 95 หน้า

งานวิจัยนี้มีวัตถุประสงค์เพื่อเปรียบเทียบวิธีการประมาณค่าข้อมูลสูญหายสำหรับข้อมูลตัวแปรพหุ  
2 วิธี คือ Expectation Maximization Algorithm (EM) และ Regularized Expectation Maximization  
Algorithm (REM) เกณฑ์ที่ใช้ในการเปรียบเทียบคือ ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) โดยที่  
ถ้าวิธีการประมาณค่าสูญหายวิธีใดให้ค่า MSE ต่ำสุด แสดงว่าวิธีนั้นมีประสิทธิภาพดีกว่า ข้อมูลที่ใช้  
ในการศึกษาครั้งนี้ผู้วิจัยได้จำลองข้อมูล ด้วยเทคนิคมอนติคาร์โล จำแนกสถานการณ์โดยกำหนด  
จำนวน ตัวแปรอิสระ (X) เป็น 3, 4, 5, 7 และตัวแปรตาม (Y) เป็น 2 และ 3 ขนาดตัวอย่าง (n) เป็น 50 70  
และ 100 และกำหนดให้ระดับความสัมพันธ์ระหว่างตัวแปรอิสระระดับสูง ( $\rho = 0.7 - 0.9$ ) และระดับต่ำ  
( $\rho = 0.1 - 0.3$ ) ตัวแปรตามมีข้อมูลสูญหายแบบสุ่ม (Missing at Random, MAR) ที่ระดับการสูญหาย  
10%, 20% และ 30% ดังนั้นข้อมูลที่จำลองตามสถานการณ์ต่าง ๆ มีจำนวนทั้งสิ้น 144 สถานการณ์ โดย  
ทำซ้ำแต่ละสถานการณ์จำนวน 1,000 ครั้ง และศึกษาวิธีประมาณค่าสูญหาย กับข้อมูลจริงจำนวน 2 ชุด  
คือ 1) ข้อมูลกลุ่มการทรงตัวของบรรยากาศและค่าพยากรณ์อากาศ และ 2) ข้อมูลกลุ่มความชื้น ของ  
ข้อมูลผลการตรวจอากาศชั้นบน ระหว่างวันที่ 1 มีนาคม 2549 - 31 ตุลาคม 2551 จากสถานีเรดาร์  
อำเภอพิมาย จังหวัดนครราชสีมา สำนักฝนหลวงและการบินเกษตร

ผลการศึกษาจากการพิจารณาค่า MSE สำหรับข้อมูลจำลอง พบว่า กรณีระดับความสัมพันธ์  
ระหว่างตัวแปรอิสระสูง และจำนวนตัวแปรอิสระ 3 และ 4 ตัวแปร ขนาดตัวอย่าง 50 ที่ทุกระดับการสูญหาย  
วิธี EM เป็นวิธีที่เหมาะสม ส่วนขนาดตัวอย่าง 70 และ 100 ที่ทุกระดับการสูญหาย วิธี REM เป็นวิธีที่  
เหมาะสม สำหรับจำนวนตัวแปรอิสระ 5 และ 7 ตัวแปร ที่ทุกขนาดตัวอย่าง และเกือบทุกระดับการสูญ  
หาย วิธี REM เป็นวิธีที่เหมาะสม กรณีระดับความสัมพันธ์ระหว่างตัวแปรอิสระต่ำ และจำนวนตัวแปร  
อิสระ 3 และ 4 ตัวแปร ที่ทุกขนาดตัวอย่าง เกือบทุกระดับการสูญหาย วิธี EM เป็นวิธีที่เหมาะสม  
สำหรับจำนวนตัวแปรอิสระ 5 และ 7 ตัวแปร ที่ทุกขนาดตัวอย่าง เกือบทุกระดับการสูญหาย วิธี REM  
เป็นวิธีที่เหมาะสม ส่วนผลการศึกษากับข้อมูลจริงจำนวน 2 ชุด พบว่า วิธี REM เป็นวิธีที่เหมาะสม ซึ่ง  
ผลการศึกษาล้วนใหญ่สอดคล้องกับผลการศึกษาจากข้อมูลจำลอง