3837242 PHBS/M : สาขาวิชา : ชีวสถิติ; วท.ม. (ชีวสถิติ)
คำสำคัญ : การใส่ค่าหลายค่าแทนข้อมูลที่สูญหายแต่ละค่า / ข้อมูลสูญหาย

บวรวรรณ ดิเรกโภค : การประยุกต์วิธีการใส่ค่าหลายค่าแทนข้อมูลที่สูญหายแต่ละค่าในการวิเคราะห์ข้อมูลอุบัติเหตุผู้ขับขี่จักรยานยนต์ (AN APPLICATION OF MULTIPLE IMPUTATION FOR ANALYSING ACCIDENT DATA OF MOTORCYCLE DRIVERS) คณะกรรมการควบคุมวิทยานิพนธ์ : ธวัชชัย วรพงศธร, Ph.D., ชูเกียรติ วิวัฒน์วงศ์เกษม, วท.ม. 106 หน้า. ISBN 974-664-180-8

ข้อมูลที่สูญหายจะลดประสิทธิภาพของค่าประมาณจากการวิเคราะห์ฐานข้อมูลที่เหลืออยู่ และทำให้ไม่สามารถวิเคราะห์ข้อมูลด้วยวิธีมาตรฐานได้ตามปกติ วิธีใส่ค่าหลายค่าแทนข้อมูลที่สูญหายจะทำให้ค่าประมาณทางสถิติมีความถูกต้องกว่าวิธีใส่ค่าเพียงค่าเดียว

การศึกษานี้มีวัตถุประสงค์เพื่อเปรียบเทียบค่าความคลาดเคลื่อนของสัมประสิทธิ์การถดถอยแบบลอจิสติคระหว่างวิธีการใส่ค่าหลายค่าแทนข้อมูลที่สูญหายแต่ละค่าแบบสุ่มอย่างง่าย เมื่อตัวแปรมีข้อมูลสูญหาย โดยใช้จำนวนค่าที่ใส่แทนค่าสูญหายแต่ละค่า M=3, 5 กับการใส่ค่าเพียงค่าเดียว M=1 โดยใช้ข้อตกลงเบื้องต้นว่าข้อมูลมีการสูญหายแบบเชิงสุ่ม (MAR) ตัวอย่างการศึกษาคือข้อมูลทุติยภูมิผู้ขับขี่จักรยานยนต์ที่บาดเจ็บเข้ารับการรักษาที่แผนกฉุกเฉิน โรงพยาบาลราชวิถี จำนวน 2,668 ราย ตั้งแต่วันที่ 1 มกราคม - 31 ธันวาคม 2538

ผลของการศึกษาด้วยการประยุกต์ใช้วิธีการใส่ค่าหลายค่าแทนค่าสูญหายแต่ละค่า ชี้ให้เห็นว่าค่าความคลาดเคลื่อนของสัมประสิทธิ์การถดถอยแบบลอจิสติคเมื่อใช้ M=3, 5 มีค่ามากกว่าการใส่ค่าแทนข้อมูลที่สูญหายเพียงค่าเดียว M=1 ในโมเดลการถดถอยแบบลอจิสติคซึ่งมีตัวแปรตาม คือ ระดับคะแนนโคม่าปกติ ส่วนตัวแปรอิสระ ได้แก่ การใช้ยาที่มีผลต่อการควบคุมพาหนะขณะขับขี่, การใช้หมวกนิรภัย, การดื่มเครื่องดื่มผสมแอลกอฮอล์ของผู้ขับขี่, อายุ และ เพศ ซึ่งตัวแปรสามตัวแรกเป็นตัวแปรร่วมทวิที่มีค่าสูญหาย เมื่อเปรียบเทียบค่าประมาณในระหว่างชุดข้อมูลที่ใส่ค่าหลายค่าพบว่าค่าความคลาดเคลื่อนของสัมประสิทธิ์การถดถอยแบบลอจิสติคของตัวแปรอิสระ เมื่อใช้ M=3 มีค่ามากกว่า M=5 ยกเว้นตัวแปรการใช้หมวกนิรภัย และความคลาดเคลื่อนของสัมประสิทธิ์การถดถอยแบบลอจิสติคจากการวิเคราะห์ชุดข้อมูลที่สมบูรณ์มีค่าสูงกว่าข้อมูลที่ใส่ค่าแทนข้อมูลที่สูญหายแต่ละค่า จากการวิเคราะห์ชุดข้อมูลที่ใส่ค่าหลายค่าแทนข้อมูลที่สูญหายแต่ละค่าแสดงให้เห็นว่าโมเดลการถดถอยแบบลอจิสติคมีนัยสำคัญทางสถิติ วิธีนี้ดีกว่าการวิเคราะห์ข้อมูลที่มีเพียงหน่วยตัวอย่างซึ่งได้ค่าสังเกตของตัวแปรครบทุกตัวเท่านั้น และวิธีนี้สามารถประยุกต์ใช้ได้กับข้อมูลที่ตัวแปรไม่ได้รับคำตอบ ซึ่งจะได้ผลดีกับตัวแปรชนิดแบ่งกลุ่ม