

การบ้านที่ 10

Intro to MLM

การทดสอบทั้งหมดใช้ $\alpha = .05$ หรือช่วงเชื่อมั่นระดับ .95

ในโรงภาพยนตร์แห่งหนึ่ง ได้เก็บการให้คะแนนภาพยนตร์ของลูกค้าหลังจากที่ดูภาพยนตร์ต่างๆ จำนวน 100 เรื่อง โดยการเก็บข้อมูลลูกค้าที่จะเป็นการเก็บแบบง่ายๆ จากการถามหลังจากที่ดูภาพยนตร์เสร็จ ว่าให้คะแนนเท่าไรจาก 1 (ไม่ชอบอย่างยิ่ง) ถึง 10 (ชอบอย่างยิ่ง) แล้วผู้ถาม จะกรอกข้อมูลเพศและอายุจากการประเมิน ลงในสมุดจดของตัวเอง หากลูกค้าที่ถามไม่ประสงค์จะให้คะแนน จะบันทึกเป็นค่าสูญหาย รวมถึงหากผู้ถาม ไม่สามารถประเมินอายุของผู้ตอบได้ จะบันทึกเป็นค่าสูญหายเช่นกัน

จากการเก็บข้อมูลลูกค้าจำนวน 2885 คน ได้ข้อมูลตัวแปรดังต่อไปนี้

custid = ID ของลูกค้า

movieid = ID ของภาพยนตร์ที่ดู

score = คะแนนที่ประเมิน มีค่าตั้งแต่ 1 ถึง 10 โดย 99 คือค่าสูญหาย

age = อายุของลูกค้า โดย 99 คือค่าสูญหาย

female = เพศของลูกค้า โดย 1 = หญิง และ 0 = ชาย

animation = ประเภทของภาพยนตร์ โดย 1 = ภาพยนตร์แอนิเมชัน และ 0 = ภาพยนตร์อื่นๆ

ข้อมูลนี้อยู่ในไฟล์ที่ชื่อ homework10.csv ขอให้ท่านใช้การแทนค่าสูญหายแบบพหุ (Multiple Imputation) ที่เหมาะสม โดยจำนวนชุดข้อมูลขั้นต่ำเท่ากับ 20 ชุด และวิเคราะห์ว่าตัวแปรอะไรที่มีอิทธิพลต่อคะแนนที่ประเมินของลูกค้า และมีปฏิสัมพันธ์ระหว่างตัวแปรหรือไม่