

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อ (1) เพื่อปรับปรุงวิธีการตรวจหาค่านอกกลุ่มแบบหลายขั้นตอนของ Marasinghe (1985) ให้เป็นวิธีหลายขั้นตอนแบบสองเฟส เพื่อให้สามารถตรวจสอบค่านอกกลุ่มในการถอดออกอย่างง่ายทั้งในตัวแปรตามและตัวแปรอิสระ โดยเฟส I เป็นการหาเซตรายการค่าสังเกตซึ่งสงสัยว่าจะเป็นค่านอกกลุ่ม โดยใช้วิธีหลายขั้นตอนกับเซตค่านอกกลุ่มในตัวแปรตาม และใช้ Cook's distance สำหรับตัวแปรอิสระ เฟส II เป็นการทดสอบสมาชิกในรายการค่าสังเกตซึ่งสงสัยว่าจะเป็นค่านอกกลุ่มว่าค่าใดเป็นค่านอกกลุ่ม โดยใช้ค่าของตัวสถิติ $GESR(\max |r_i|)$ (2) เพื่อเปรียบเทียบเปอร์เซ็นต์ความถูกต้องในการตรวจพบค่านอกกลุ่มระหว่างวิธีหลายขั้นตอนแบบสองเฟสกับวิธี GESR แบบสองเฟส (Paul and Fung 1991) ในการศึกษาใช้การจำลองแบบตัวแบบการถอดออกอย่างง่าย โดยศึกษาที่สัมประสิทธิ์การถอดออก (β_1) เท่ากับ 1, 5 และ 10 ขนาดตัวอย่าง 15, 25 และ 50 สร้างค่านอกกลุ่มโดยบวกค่า δ_1 และ $\delta_2(0, 5, -5, 9$ หรือ $-9)$ ในสองค่าใดๆ ของตัวแปรตามอย่างสุ่ม และบวกค่า δ_3 และ $\delta_4(0, 2, -2, 4$ หรือ $-4)$ ในสองค่าใดๆ ของตัวแปรอิสระอย่างสุ่ม แต่ละการทดลองทำซ้ำ 1,000 ครั้ง

ผลการวิจัยพบว่า

(1) ส่วนใหญ่วิธีหลายขั้นตอนแบบสองเฟส ให้เปอร์เซ็นต์ความถูกต้องในการตรวจพบค่านอกกลุ่มสูงกว่าวิธี GESR แบบสองเฟส เล็กน้อย

(2) เมื่อ $\beta_1 = 5$ และ 10 ทั้งสองวิธีสามารถตรวจพบค่านอกกลุ่มทั้งในตัวแปรตามและตัวแปรอิสระได้ดี (มากกว่า 90%) เมื่อ $\beta_1 = 1$ ทั้งสองวิธีตรวจพบค่านอกกลุ่มในตัวแปรอิสระได้ไม่คืบคืบ แต่เมื่อขนาดตัวอย่างเพิ่มขึ้นทั้งสองวิธีให้เปอร์เซ็นต์ความถูกต้องในการตรวจพบค่านอกกลุ่มเพิ่มขึ้นอย่างเห็นได้ชัด

(3) สำหรับกรณีที่ค่านอกกลุ่มอยู่ด้านเคียวกันหรืออยู่คนละด้านของค่าเฉลี่ย ระยะห่างของค่านอกกลุ่มกับค่าเฉลี่ย ไม่ค่อยส่งผลต่อเปอร์เซ็นต์ความถูกต้องในการตรวจพบค่านอกกลุ่มมากนัก ยกเว้นเมื่อ $\beta_1 = 1$ ทั้งสองวิธีให้เปอร์เซ็นต์ความถูกต้องในการตรวจพบค่านอกกลุ่มกรณีที่ค่านอกกลุ่มในตัวแปรอิสระที่อยู่ห่างจากค่าเฉลี่ยสูงกว่ากรณีที่ค่านอกกลุ่มอยู่ใกล้ค่าเฉลี่ย

The objectives of this research were :

(1) To modify the Multistage procedure of Marasinghe (1985) to be Two-phase Multistage procedure that could detect outliers either in response variable or in independent variables. In phase I, suspected outliers were identified using the Multistage procedure for response variable and the Cook's distance for independent variables. In phase II, each observation in the suspected set was tested for its outlyingness using GESR ($\max |r_i|$) statistic.

(2) To compare percentages of exactitude in outliers detection between Two-phase GESR (Paul and Fung 1991) and Two-phase Multistage procedure.

The simulation study based on simple linear regression were conducted. The regression coefficient (β_1) was set at 1, 5 and 10. The sample size (n) was set at 15, 25 and 50. The outliers were introduced by adding known value $\delta_1, \delta_2 (0, 5, -5, 9 \text{ or } -9)$ to two values of response variable and $\delta_3, \delta_4 (0, 2, -2, 4 \text{ or } -4)$ to two values of independent variable, randomly. Each experiment has 1000 replications.

The results of the study were :

(1) The Two-phase Multistage procedure often gives slightly higher percentages of exactitude in outliers detection than the Two-phase GESR procedure.

(2) When $\beta_1 = 5$ and 10, both procedures are able to detect outliers in both response variable and independent variable very well (more than 90%). When $\beta_1 = 1$, both procedures fail to detect outliers in independent variable. However, when sample sizes increase, the percentages of outliers detection increase, obviously.

(3) The impact of the position of outliers, in the same side or different side of mean and their distances to the mean are not obvious. Except when $\beta_1 = 1$, outliers in independent variable were detected with large distance from the mean give higher percentages of exactitude in outliers detection than those with the small distance from the mean.