

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของวิธีการตรวจสอบอิทธิพลในสมการถดถอยเชิงพหุตัวแปรอิสระ 2 และ 3 ตัว และสมการถดถอยโลจิสติก ระหว่างสถิติพีนา กับ สถิติค็อก ขนาดตัวอย่างเป็น 6, 10, 20 และ 40 กำหนดจำนวนข้อมูลที่มีอิทธิพลเป็น 1, 2, 3 และ 4 ค่า ตามลำดับ ขนาดของความคลาดเคลื่อนที่ผิดปกติเป็น 6, 9, 12 เท่าของส่วนเบี่ยงเบนมาตรฐาน การแจกแจงของค่าคลาดเคลื่อนของตัวแบบกำหนดไว้ 3 แบบคือ แบบปกติ  $N(0, 1)$  การแจกแจงแบบยูนิฟอร์ม  $U(0, 2)$  และการแจกแจงแบบลาปลาซ  $La(0, \frac{1}{\sqrt{2}})$  วิธีการศึกษาเป็นการจำลองค่าข้อมูลโดยโปรแกรมสำเร็จรูป MATLAB ดำเนินการทำซ้ำ 100 ครั้ง ในทุก ๆ กรณี เกณฑ์ที่กำหนดเพื่อเปรียบเทียบประสิทธิภาพระหว่างวิธีการคือ ค่าเปอร์เซ็นต์ความถูกต้องในการบ่งชี้ข้อมูลอิทธิพลได้ตรงและครบจำนวน

ผลการศึกษาพบว่า วิธีการตรวจสอบอิทธิพลของสถิติค็อกสามารถบ่งชี้ว่ามีข้อมูลผิดปกติได้ตรงตำแหน่งและครบจำนวนตามที่กำหนดขึ้นในการจำลองแบบได้ครบ 100 เปอร์เซ็นต์แทบทุกกรณีซึ่งเปรียบเทียบแล้วดีกว่าวิธีการตรวจสอบอิทธิพลของสถิติพีนาอย่างเด่นชัด ส่วนวิธีการของสถิติพีนาถึงแม้จะไม่เหมาะสมในแง่ที่มีประสิทธิภาพต่ำกว่าสถิติค็อกหากนำสถิติพีนาไปใช้ตรวจสอบอิทธิพลของข้อมูลในสมการโลจิสติกก็ยิ่งทำให้มีประสิทธิภาพต่ำกว่าการใช้ในสมการถดถอยเชิงเส้นพหุ

The objective of this research is to compare the efficiency of detecting the influence data between Pena's statistic and Cook's statistic for the multiple linear regression models with 2 and 3 independent variables and logistic regression model.

The sample size of 6, 10, 20, and 40 is set, the number of influence data is set to be 1, 2, 3, and 4. The size of outliers are set in terms of 6, 9, and 12 times of the standard deviation of the error term of the models. The distributions of the error, as the factor of the study, are  $N(0, 1)$ ,  $U(0, 2)$ , and  $La(0, \frac{1}{\sqrt{2}})$ . The simulation of data and analysis is done by MATLAB software programming for 100 times replication in each case. The comparison criteria is the corrected percentage which is the ability to identify exact and all of influence data in the model correctly.

The results of the study are as followed : the detecting method by Cook's statistic is able to identify the influence data exactly and totally 100 percent in almost cases of study which is much more effective than the Pena's statistic. Even through the Pena's method is less effective than the Cook's statistic, if we use Pena's detecting method in logistic regression the corrected percentage is lesser than the multiple linear regression.