การเปรียบเทียบประสิทธิภาพของสถิติทดสอบสำหรับการแจกแจงปกติ ชุดคำสั่งนอร์เทสต์ในโปรแกรมอาร์

AN EFFICIENCY COMPARISON OF TEST STATISTIC FOR NORMAL DISTRIBUTION OF NORTEST PACKAGE IN R PROGRAM

ศุภรัตน์ มีประพันธ์[1]* *อัชฌา อระวีพร*[2]

Suparat Meepraphan[1] ***,** *Autcha Araveeporn*[2]

[1]*ภาควิชาสถิติ คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง*

[1]*Department of Statistics, Faculty of Science, King Mongkut's Institute of Technology Ladkrabang.*

[2]*คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง*

[2]*Faculty of Science, King Mongkut's Institute of Technology Ladkrabang.*

***Corresponding author, E-mail:** *tarn_math@hotmail.com*

บทคัดย่อ

การวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาเปรียบเทียบความสามารถในการควบคุมความคลาดเคลื่อนชนิดที่ 1 และอำนาจของการทดสอบของสถิติทดสอบสำหรับการแจกแจงปกติ 6 การทดสอบของชุดคำสั่งนอร์เทสต์ (Nortest) ในโปรแกรมอาร์ (R) ได้แก่ สถิติทดสอบ Kolmogorov-Sminov (KS) สถิติทดสอบ Shapiro-Francia (SF) สถิติทดสอบ Lilliefor (LF) สถิติทดสอบ Anderson-darling (AD) สถิติทดสอบ Cramer-von Mises (CVM) และ สถิติทดสอบ Pearson chi-square (PCS) โดยศึกษาจากประชากรที่มีการแจกแจงปกติที่ค่าเฉลี่ย 0 ความแปรปรวน 1 และ 4 (N(0,1))(N(0,4)) และ (N(0,8)) การแจกแจงทีที่องศาเสรี 4(t_4)12 (t_{12}) และ 100 (t_{100}) การแจกแจงแกมมาที่ค่าพารามิเตอร์แสดงถึงรูปร่าง 24 และ 1 (α) และที่ค่าพารามิเตอร์แสดงถึงสเกล 41/2 และ 1/3(β)(Gamma(2,4) Gamma(4,1/2) และ Gamma(1,1/3)) การแจกแจงทวินามที่ค่าพารามิเตอร์ (p) เท่ากับ 0.2 0.3 0.5 0.6 และ 0.7 การแจกแจงปัวซงที่ค่าพารามิเตอร์ (λ) เท่ากับ 0.5 5 10 และ 20 กำหนดขนาดตัวอย่างเท่ากับ 20 30 50 70 และ 100 กำหนดระดับนัยสำคัญ 3 ระดับ คือ 0.01 0.05 และ 0.1 ใช้โปรแกรมอาร์ (R) เวอร์ชัน 3.0.2 ในการจำลองและวิเคราะห์ข้อมูลทำซ้ำ 5,000 รอบในแต่ละสถานการณ์

ผลการวิจัยพบว่า ในกรณีที่ลักษณะของข้อมูลมีการแจกแจงปกติ N(0,1) N(0,4) และ N(0,8) สถิติทดสอบ KS สถิติทดสอบ LF และ สถิติทดสอบ CVM สามารถควบคุมความคลาดเคลื่อนชนิดที่ 1 ได้ทั้งหมดที่ทุกขนาดตัวอย่างและระดับนัยสำคัญ การแจกแจงที การแจกแจงแกมมา Gamma (2,0.25) สถิติทดสอบ SF มีอำนาจการทดสอบสูงที่สุดทุกขนาดตัวอย่างและระดับนัยสำคัญ

คำสำคัญ: สถิติทดสอบ Shapiro-Francia สถิติทดสอบ Lilliefor สถิติทดสอบ Anderson-darling สถิติทดสอบ Cramer-von Mises สถิติทดสอบ Pearson chi-square ความคลาดเคลื่อนชนิดที่ 1 อำนาจการทดสอบ

Abstract

The objective of this research is to compare the performance for controlling the type I error and the power of the test from test statisticofnormal distribution via program Rin case of Nortest package. The test statistics incase of Nortest package consist of Shapiro-Francia (SF) test, Lilliefor (LF) test, Anderson-Darling (AD) test, Cramer-von Mises (CVM) test, and Pearson Chi-square (PCS) test. In this case, data is generated from normal distribution with mean 0 and variance 1 and 4 or called N(0,1) and N(0,4), T distribution with degree of freedom 4 or called , and gamma distribution with shape parameter 2 and scale parameter 0.25 or called Gamma (2,0.25). The sample size isset at 20, 30, 50 and 70and significant level is specified at 0.01 and 0.05. The program R version 3.0.2 is used bygenerating dataand dataanalysis with 5000 replicationsin each situation.

According to the result,normal distribution at N(0,1) and N(0,4), SF test, LF test, AD test, CVM test, and PCS test can control type I error at all sample sizes and significant levels. For T distribution at and gamma distribution at Gamma (2, 0.25), Shapiro-Francia (SF) test presents the highest of power of the test for all sample sizes and significant levels.

Keywords: Shapiro-Francia test, Lilliefor test, Anderson-darling test, Cramer-von Mises test, Pearson Chi-square test, Type I error, Power of the test

บทนำ

การศึกษาลักษณะของประชากร (Population) ทั้งหมดเป็นไปได้ยาก จึงนิยมใช้การสุ่มหาตัวแทนของประชากรหรือเรียกว่าตัวอย่าง (Sample) เป็นตัวแทนในการศึกษา โดยทั่วไปเพื่อต้องการทราบลักษณะของประชากรทั้งหมด จึงต้องมีการอนุมานข้อมูลจากตัวอย่างไปหาข้อมูลจากประชากร ซึ่งคือสถิติเชิงอนุมาน (Inferential Statistics) มีการนำทฤษฎีความน่าจะเป็น (Probability Theory) เข้ามาประยุกต์ใช้ เช่น การทดสอบสมมติฐาน (Hypothesis Testing)

การทดสอบสมมุติฐานเพื่อเปรียบเทียบความแตกต่างของค่าเฉลี่ยของประชากร 1 และ 2 ประชากรที่เป็นอิสระกันจะใช้สถิติทดสอบ t เมื่อขนาดตัวอย่างน้อยกว่า 30 และไม่ทราบค่าความแปรปรวนของประชากรและสถิติทดสอบ Z เมื่อทราบค่าความแปรปรวนของประชากรและถ้าต้องการทดสอบค่าเฉลี่ยของประชากรมากกว่า 2 กลุ่มและความแปรปรวนของประชากรเท่ากันจะใช้สถิติทดสอบ F ในสถิติทดสอบที่กล่าวมานี้มีเงื่อนไขสำคัญคือตัวอย่างที่ได้ต้องมาจากการสุ่มจากประชากรที่มีการแจกแจงปกติ

โปรแกรมอาร์ (R) มีชุดคำสั่งสำหรับตรวจสอบข้อมูลของการแจกแจงปกติ อยู่ในชุดคำสั่งที่เรียกว่า นอร์เทสต์ (Nortest) ซึ่งเป็นโปรแกรมที่ไม่มีลิขสิทธิ์สามารถดาวน์โหลดโดยไม่เสียค่าใช้จ่ายและสามารถวิเคราะห์ข้อมูลทางสถิติได้หลากหลาย ดังนี้สามารถใช้จำลองข้อมูลพร้อมทั้งวิเคราะห์ตามความต้องการของผู้ใช้ได้อีกด้วย

ในการทำวิจัย จึงสนใจทำการเปรียบเทียบความสามารถในการควบคุมความคลาดเคลื่อนชนิดที่ 1 และอำนาจการทดสอบโดยใช้โปรแกรมอาร์ (R) เวอร์ชัน 3.0.2 สำหรับการจำลองและวิเคราะห์ข้อมูลจากชุดคำสั่งนอร์เทสต์ (Nortest) ประกอบด้วย

สถิติทดสอบ Shapiro-Francia (SF) สถิติทดสอบ Lilliefor (LF) สถิติทดสอบ Anderson-darling (AD) สถิติทดสอบ Cramer-von Mises (CVM) และสถิติทดสอบ Pearson chi-square (PCS) เนื่องจาก 5 การทดสอบดังกล่าวศึกษารูปแบบของฟังก์ชันการแจกแจงสะสมของตัวอย่างในการทดสอบ

วัตถุประสงค์ของการวิจัย

1. เพื่อเปรียบเทียบความสามารถในการควบคุมความคลาดเคลื่อนชนิดที่ 1 และอำนาจการทดสอบของสถิติทดสอบสำหรับการแจกแจงปกติ 5 การทดสอบ คือ สถิติทดสอบ SF สถิติทดสอบ LF สถิติทดสอบ AD สถิติทดสอบ CVM และสถิติทดสอบ PCS

2. เพื่อศึกษาสถิติทดสอบที่เหมาะสมสำหรับทดสอบข้อมูลที่มีการแจกแจงปกติ การแจกแจงที การแจกแจงแกมมา

วิธีดำเนินการวิจัย

การวิจัยนี้เป็นการวิจัยเชิงทดลองเพื่อเปรียบเทียบประสิทธิภาพของสถิติทดสอบสำหรับการแจกแจงปกติของชุดคำสั่งนอร์เทสต์ในโปรแกรมอาร์ (R) ที่ขนาดตัวอย่างเท่ากับ 20 30 50 และ 70 และระดับนัยสำคัญ 2 ระดับ คือ 0.01 และ 0.05

1. ขั้นตอนในการคำนวณความสามารถในการควบคุมความคลาดเคลื่อนชนิดที่ 1(α)

1.1 จำลองข้อมูลในแต่ละขนาดจากประชากรที่มีการแจกแจงปกติที่ค่าเฉลี่ย 0 ความแปรปรวน 1 (N(0,1)) และการแจกแจงปกติที่ค่าเฉลี่ย 0 ความแปรปรวน 4 (N(0,4)) โดยโปรแกรมอาร์ (R)

1.2 คำนวณสถิติทดสอบทั้ง 5 การทดสอบ ได้แก่ สถิติทดสอบ SF สถิติทดสอบ LF สถิติทดสอบ AD สถิติทดสอบ CVM และสถิติทดสอบ PCS โดยใช้ชุดคำสั่งนอร์เทสต์ในโปรแกรมอาร์ (R)

1.3 ทำการทดสอบสมมุติฐานจากสถิติทดสอบทั้ง 5 การทดสอบ โดยนำค่า p-value มาเปรียบเทียบกับระดับนัยสำคัญ โดยจะปฏิเสธสมมติฐานว่างถ้าค่า p-value น้อยกว่าระดับนัยสำคัญทั้ง 2 ระดับ

1.4 ทำซ้ำข้อ 1.1 –1.3 จนครบ 5,000 ครั้ง แล้วทำการหาความคลาดเคลื่อนชนิดที่ 1 โดยการนับจำนวนครั้งของการปฏิเสธสมมติฐานว่าง (H_0) ดังนี้

$$\text{ความคลาดเคลื่อนชนิดที่ 1} = \frac{\text{จำนวนครั้งของการปฏิเสธ } H_0 \text{ เมื่อ } H_0 \text{ เป็นจริง}}{5{,}000}$$

ถ้าความสามารถในการควบคุมความคลาดเคลื่อนชนิดที่ 1 ของการทดสอบสำหรับแต่ละสถานการณ์มีค่าอยู่ในช่วงที่ได้กำหนดไว้ในเกณฑ์ของการเปรียบเทียบวิธีการทดสอบ ได้แก่ เกณฑ์ของ Bradley (1978) [1] และเกณฑ์ของ Cochran (1954) [2] จะถือว่าสถิติทดสอบนั้นมีความสามารถในการควบคุมความคลาดเคลื่อนชนิดที่ 1 ได้เกณฑ์ของ Bradley (1978) [1] กำหนดไว้ดังนี้

การทดสอบที่ระดับนัยสำคัญ 0.01 ความคลาดเคลื่อนชนิดที่ 1 จากการทดลองอยู่ในช่วง (0.005, 0.015)

การทดสอบที่ระดับนัยสำคัญ 0.05 ความคลาดเคลื่อนชนิดที่ 1 จากการทดลองอยู่ในช่วง (0.025, 0.075)

เกณฑ์ของ Cochran (1954) [2] กำหนดไว้ดังนี้

การทดสอบที่ระดับนัยสำคัญ 0.01 ความคลาดเคลื่อนชนิดที่ 1 จากการทดลองอยู่ในช่วง (0.007, 0.015)

การทดสอบที่ระดับนัยสำคัญ 0.05 ความคลาดเคลื่อนชนิดที่ 1 จากการทดลองอยู่ในช่วง (0.040, 0.060)

2. ขั้นตอนในการคำนวณอำนาจการทดสอบของสถิติทดสอบ (1-β)

2.1 จำลองข้อมูลในแต่ละขนาดจากประชากรที่มีการแจกแจงที ที่องศาเสรี 4(t_4) และการแจกแจงแกมมาที่ค่าพารามิเตอร์แสดงถึงรูปร่าง 2 (α) และที่ค่าพารามิเตอร์แสดงถึงสเกล 0.25 (β) (Gamma (2, 0.25) ด้วยโปรแกรมอาร์ (R)

2.2 คำนวณสถิติทดสอบทั้ง 5 การทดสอบ ได้แก่ สถิติทดสอบ SF สถิติทดสอบ LF สถิติทดสอบ AD สถิติทดสอบ CVM และสถิติทดสอบ PCS โดยใช้ชุดคำสั่งนอร์เทสต์ในโปรแกรมอาร์ (R)

2.3 ทำการทดสอบสมมติฐานจากสถิติทดสอบทั้ง 5 การทดสอบ โดยนำค่า p-value มาเปรียบเทียบกับระดับนัยสำคัญ โดยจะปฏิเสธสมมติฐานว่างถ้าค่า p-value น้อยกว่าระดับนัยสำคัญทั้ง 2 ระดับ

2.4 ทำซ้ำข้อ 2.1 -2.3 จบครบ 5,000 ครั้ง แล้วหาอำนาจการทดสอบ โดยการนับจำนวนครั้งของการปฏิเสธสมมติฐานว่าง (H_0) ดังนี้

$$\text{อำนาจการทดสอบ} = \frac{\text{จำนวนครั้งของการปฏิเสธ } H_0 \text{ เมื่อ } H_0 \text{ ไม่เป็นจริง}}{5{,}000}$$

3. สถิติทดสอบของการแจกแจงปกติ

การทดสอบการแจกแจงปกติ สมมติฐานที่ต้องการทดสอบ คือ

H_0 : ประชากรมีการแจกแจงปกติ

H_1 : ประชากรไม่มีการแจกแจงปกติ

3.1 สถิติทดสอบ Shapiro–Francia (SF)

มีหลักการเดียวกับสถิติทดสอบ Shapiro และ Wilk ใช้ค่าคาดหวังของสถิติอันดับของการแจกแจงปรกติ (Normal ordered statistics) เข้าช่วยการคำนวณทำให้เข้าใจได้ง่ายขึ้น ขนาดตัวอย่างที่ใช้ตั้งแต่ 35 ถึง 99 ภายหลัง Royston (1993) [3] ได้ปรับตัวสถิติทดสอบ Shapiro และ Francia เพื่อให้ใช้ได้กับขนาดตัวอย่างที่กว้างขึ้น [4-5]

สถิติที่ใช้ทดสอบ

$$SF = \frac{\left(\sum_{i=1}^{n} m_i x_i\right)^2}{\left(\sum_{i=1}^{n} m_i^2 \times \sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2\right)}$$

$$y = \frac{\left(1-SF\right)^{\lambda} - 1}{\lambda} \quad \text{และ} \quad Z = \frac{\left(y-\mu_y\right)}{\sigma_y}$$

เมื่อ

Z คือ ค่าปกติมาตรฐาน

μ_y คือ ค่าเฉลี่ยของ y

σ_y คือ ค่าส่วนเบี่ยงเบนมาตรฐานของ y

$\bar{x}$ คือ ค่าเฉลี่ยเลขคณิต

x_i คือ ค่าของข้อมูลตัวที่ i

m_i คือ ค่าจากตาราง Expected values of normal order statistics

ค่าประมาณ λ μ_y และ σ_y คำนวณได้ดังนี้

$$\lambda = -0.048157 + 0.0197196x - 0.0119065x^3$$

$$\mu_y = -exp\left(1.693067 + 0.144167x - 0.0184928x^2 + 0.031074485x^3 + 0.0055717663x^4\right)$$

$$\sigma_y = exp\left(-0.510725 - 0.1160364x - 0.0067021x^2 + 0.054465944x^3 + 0.0087397329x^4\right)$$

$$x = \log\left(n\right) - 5$$

ตัวเลขที่ใช้ในการหาค่าต่างๆ เกิดจากการสร้างตัวแบบแล้วจึงแปลงข้อมูลด้วยวิธีบอกซ์-คอกซ์ (Box-Cox Transformation) ทำให้การประมาณค่าให้เข้าใกล้การแจกแจงปรกติของตัวอย่างสุ่ม [5]

การตัดสินใจปฏิเสธสมมติฐานเมื่อค่า SF ที่คำนวณได้มีค่ามากกว่าค่า $Z = \dfrac{\left(y-\mu_y\right)}{\sigma_y}$

3.2 สถิติทดสอบ Lilliefor (LF)

ใช้ฟังก์ชันการแจกแจงสะสมของตัวอย่าง (Empirical Distribution Function) ประกอบด้วยสมมติฐานของการแจกแจงเป็นปรกติ พิจารณาที่ค่าความแตกต่างสูงสุดระหว่างฟังก์ชันเชิงประจักษ์ (Empirical function) และฟังก์ชันการแจกแจงสะสม [6]

สถิติที่ใช้ทดสอบ LF คือ $D = \max\left\{D^+, D^-\right\}$

$$D^+ = \max_{i=1,\ldots,n}\left\{\frac{1}{n} - p_i\right\}$$

$$D^- = \max_{i=1,\ldots,n}\left\{p_i - \frac{(i-1)}{n}\right\}$$

$$p_i = \Phi\left(\frac{|x_i - \bar{x}|}{s}\right) \quad ; \quad \Phi = F_0 \; \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

เมื่อ	Φ	คือฟังก์ชันการแจกแจงสะสมของการแจกแจงปกติมาตรฐาน
	$\bar{x}$	คือค่าเฉลี่ยเลขคณิต
	x_i	คือค่าของข้อมูลตัวที่ i
	s	คือค่าส่วนเบี่ยงเบนมาตรฐาน
	n	คือจำนวนค่าสังเกต

สำหรับทุกค่าของ $i = 1, 2, 3, \ldots, n$

การตัดสินใจปฏิเสธสมมติฐานเมื่อค่า D ที่คำนวณได้มีค่ามากกว่าค่า Z ที่ได้จาก

$$Z = D\left(\sqrt{n} - 0.01 + \frac{0.85}{\sqrt{n}}\right)$$

3.3 สถิติทดสอบ Anderson-Darling (AD)

ข้อมูลอยู่ในสเกลอันดับ (Ordinal Scale) หรือลักษณะการแจกแจงของข้อมูลเป็นแบบต่อเนื่อง [7]

สถิติที่ใช้ทดสอบ $$AD = -n - \frac{1}{n}\sum_{i=1}^{n}\left[2i-1\right]\left[\ln p_i + \ln\left(1 - p_{n-i+1}\right)\right]$$

$$p_i = \Phi\left(\frac{|x_i - \bar{x}|}{s}\right) \quad ; \quad \Phi = F_0 \; \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \; s = \sqrt{\frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}{n-1}}$$

เมื่อ Φ คือฟังก์ชันการแจกแจงสะสมของการแจกแจงปกติมาตรฐาน

$\bar{x}$ คือค่าเฉลี่ยเลขคณิต

x_i คือค่าของข้อมูลตัวที่ i

s คือค่าส่วนเบี่ยงเบนมาตรฐาน

n คือจำนวนค่าสังเกต

สำหรับทุกค่าของ $i = 1, 2, 3, ..., n$

การตัดสินใจปฏิเสธสมมุติฐานเมื่อค่า AD ที่คำนวณได้มีค่ามากกว่าค่า Z ที่ได้จาก

$$Z = AD\left(1.0 + \frac{0.75}{n} + \frac{2.25}{n^2}\right)$$

3.4 สถิติทดสอบ Cramer-von Mises (CVM)

ใช้หลักการประมาณระยะห่างที่น้อยที่สุดของฟังก์ชันการแจกแจงสะสมตามการแจกแจงที่คาดไว้ เมื่อนำไปเปรียบเทียบกับฟังก์ชันการแจกแจงความถี่สัมพัทธ์สะสมของตัวอย่าง [4]

สถิติที่ใช้ทดสอบ $$CVM = \frac{1}{12n} + \sum_{i=1}^{n}\left(p_i - \frac{2i-1}{2n}\right)^2$$

$$p_i = \Phi\left(\frac{\left|x_i - \bar{x}\right|}{s}\right) \;\; ; \;\; \Phi = F_0 \;\; \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \;\; s = \sqrt{\frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}{n-1}}$$

เมื่อ Φ คือฟังก์ชันการแจกแจงสะสมของการแจกแจงปรกติมาตรฐาน

$\bar{x}$ คือค่าเฉลี่ยเลขคณิต

x_i คือค่าของข้อมูลตัวที่ i

s คือค่าส่วนเบี่ยงเบนมาตรฐาน

n คือจำนวนค่าสังเกต

สำหรับทุกค่าของ $i = 1, 2, 3, ..., n$

การตัดสินใจปฏิเสธสมมติฐานเมื่อค่า CVM ที่คำนวณได้มีค่ามากกว่าค่า Z ที่ได้จาก

$$Z = CVM\left(1.0 + \frac{0.5}{n}\right)$$

3.5 สถิติทดสอบ Pearson Chi-square (PCS)

ต้องจัดข้อมูลตัวอย่างขนาด N ออกเป็นกลุ่มย่อยๆ และบันทึกความถี่ (Frequency) ที่เกิดขึ้นในแต่ละกลุ่ม กรณีที่ข้อมูลเป็นข้อมูลเชิงคุณภาพการจัดกลุ่มอาจจัดได้ตามลักษณะข้อมูล ซึ่งอาจเป็นตัวเลขหรือไม่เป็นตัวเลขก็ได้ดังนั้นการทดสอบแบบไคกำลังสองจึงเหมาะสมกับข้อมูลที่บันทึกความถี่ในกลุ่มต่างๆ นั่นคือ การแจกแจงที่น่าจะเหมาะสมสำหรับข้อมูลนี้ คือ การแจกแจงแบบไม่ต่อเนื่อง (Discrete Distribution) การใช้การแจกแจงแบบไคกำลังสองประมาณการแจกแจงของตัวสถิติทดสอบ χ^2 จะดีเมื่อ N มีขนาดใหญ่ และ E_i ; $i = 1, 2, 3, ..., k$ ควรมีค่าอย่างน้อย 5 ขึ้นไป ในกรณีที่ค่า E_i มีค่าน้อยกว่า 5 จำเป็นต้องมีการปรับให้มีค่าตั้งแต่

5 ขึ้นไป โดยรวมกลุ่มเข้าด้วยกัน ค่า E_i น้อยกว่า 1 ในกรณีที่ไม่สามารถรวมกลุ่มเข้าด้วยกันได้เนื่องจากความหมายอาจจะเปลี่ยนไป จึงมีความจำเป็นต้องเพิ่มขนาดตัวอย่าง N ให้มีขนาดใหญ่ขึ้น [8]

สถิติที่ใช้ทดสอบ $$\chi^2 = \sum_{i=1}^{k} \frac{\left(O_i - E_i\right)^2}{E_i}$$

เมื่อ k คือจำนวนกลุ่มต่างๆ ที่ข้อมูลตัวอย่างจัดเป็นกลุ่มๆ

m คือจำนวนพารามิเตอร์ที่ต้องการประมาณค่าซึ่งจะแตกต่างไปตามการแจกแจกแจงต่างๆ

O_i คือความถี่ที่สังเกตได้จากข้อมูลตัวอย่างกลุ่มที่ i, $i = 1, 2, 3, ..., k$

E_i คือความถี่ที่คาดหวังเมื่อ H_0 เป็นจริงจากกลุ่มที่ i, $i = 1, 2, 3, ..., k$

$$\sum_{i=1}^{k} O_i = \sum_{i=1}^{k} E_i = N$$

การตัดสินใจปฏิเสธสมมติฐานเมื่อค่า PCS ที่คำนวณได้มีค่ามากกว่าค่า PCS ที่ได้จากตารางการแจกแจงแบบไคกำลังสองที่ d.f. = k - 1- m

ผลการวิจัย

การวิจัยนี้เป็นการวิจัยเชิงทดลอง เพื่อทำการศึกษาและเปรียบเทียบประสิทธิภาพของสถิติทดสอบที่ใช้การทดสอบการแจกแจงแบบปกติ 5 การทดสอบ โดยสถิติทดสอบที่เลือกมาทำการศึกษา ได้แก่ สถิติทดสอบ SF สถิติทดสอบ LF สถิติทดสอบ AD สถิติทดสอบ CVM และสถิติทดสอบ PCS โดยใช้ชุดคำสั่งนอร์เทสต์ในโปรแกรมอาร์ (R)

1. ความสามารถในการควบคุมความน่าจะเป็นของความคลาดเคลื่อนชนิดที่ 1

จากการพิจารณาความสามารถในการควบคุมความน่าจะเป็นของความคลาดเคลื่อนชนิดที่ 1 ตามเกณฑ์ของ Bradley [1] และเกณฑ์ของ Cochran [2] ที่ระดับนัยสำคัญ 0.01 สำหรับขนาดตัวอย่าง 20 30 50 และ 70 เมื่อประชากรมีการแจกแจงปกติ N(0,1) และ N(0,4) สถิติทดสอบทั้ง 4 การทดสอบสามารถควบคุมความคลาดเคลื่อนชนิดที่ 1 ได้ทั้งหมดผ่านเกณฑ์ของ Bradley [1] และเกณฑ์ของ Cochran [2] ยกเว้นเมื่อประชากรมีการแจกแจงปกติ N(0,4) ขนาดตัวอย่าง 50 สถิติทดสอบ LF ผ่านเกณฑ์ของ Bradley [1] แต่ไม่ผ่านเกณฑ์ของ Cochran [2] แสดงดังตารางที่ 1

ตารางที่ 1 ความน่าจะเป็นของความคลาดเคลื่อนชนิดที่ 1 ของสถิติทดสอบ 5 การทดสอบ ที่ระดับนัยสำคัญ 0.01

	N(0,1)					N(0,4)				
n	SF	LF	AD	CVM	PCS	SF	LF	AD	CVM	PCS
20	0.009***	0.008***	0.009***	0.009***	0.013***	0.010***	0.012***	0.010***	0.010***	0.014***
30	0.010***	0.010***	0.012***	0.012***	0.009***	0.013***	0.001***	0.011***	0.010***	0.011***
50	0.014***	0.010***	0.013***	0.013***	0.011***	0.011***	0.006*	0.009***	0.009***	0.008***
70	0.010***	0.011***	0.011***	0.012***	0.010***	0.013***	0.008***	0.008***	0.009***	0.008***

* หมายถึง ผ่านเกณฑ์ของ Bradley [1]

** หมายถึง ผ่านเกณฑ์ของ Cochran [2]

*** หมายถึง ผ่านทั้งผ่านเกณฑ์ของ Bradley [1] และเกณฑ์ของ Cochran [2]

จากการพิจารณาความสามารถในการควบคุมความน่าจะเป็นของความคลาดเคลื่อนชนิดที่ 1 ตามเกณฑ์ของ Bradley [1] และเกณฑ์ของ Cochran [2] ที่ระดับนัยสำคัญ 0.05 สำหรับขนาดตัวอย่าง 20 30 50 และ 70 เมื่อประชากรมีการแจกแจงปกติ N(0,1) และ N(0,4) สถิติทดสอบ 4 การทดสอบสามารถควบคุมความคลาดเคลื่อนชนิดที่ 1 ได้ทั้งหมดผ่านเกณฑ์ของ Bradley [1] และเกณฑ์ของ Cochran [2] ยกเว้นเมื่อประชากรมีการแจกแจงปกติ N(0,1) ขนาดตัวอย่าง 70 สถิติทดสอบ SF ผ่านเกณฑ์ของ Bradley [1] แต่ไม่ผ่านเกณฑ์ของ Cochran [2] แสดงดังตารางที่ 2

ตารางที่ 2 ความน่าจะเป็นของความคลาดเคลื่อนชนิดที่ 1 ของสถิติทดสอบ 5 การทดสอบ ที่ระดับนัยสำคัญ 0.05

	N(0,1)					N(0,4)				
n	SF	LF	AD	CVM	PCS	SF	LF	AD	CVM	PCS
20	0.051***	0.047***	0.046***	0.049***	0.051***	0.054***	0.050***	0.052***	0.051***	0.047***
30	0.051***	0.051***	0.048***	0.049***	0.049***	0.053***	0.050***	0.048***	0.047***	0.051***
50	0.050***	0.043***	0.048***	0.049***	0.052***	0.054***	0.047***	0.047***	0.048***	0.054***
70	0.061*	0.052***	0.055***	0.052***	0.051***	0.057***	0.054***	0.053***	0.051***	0.048***

* หมายถึง ผ่านเกณฑ์ของ Bradley [1]

** หมายถึง ผ่านเกณฑ์ของ Cochran [2]

*** หมายถึง ผ่านทั้งผ่านเกณฑ์ของ Bradley [1] และเกณฑ์ของ Cochran [2]

2. อำนาจการทดสอบของสถิติทดสอบ

จากการพิจารณาอำนาจการทดสอบจะใช้ข้อมูลที่ไม่ใช่การแจกแจงปกติเพื่อหาความแกร่งของสถิติทดสอบ เมื่อประชากรมีการแจกแจงที (t_4) และการแจกแจงแกมมา (Gamma(2,0.25)) ที่ระดับนัยสำคัญ 0.01 และ 0.05 สำหรับขนาดตัวอย่าง 20 30 50 และ 70 สถิติทดสอบ SF มีอำนาจการทดสอบสูงที่สุดทุกขนาดตัวอย่างและเมื่อขนาดตัวอย่างเพิ่มขึ้นอำนาจการทดสอบจะมีค่าเพิ่มขึ้นแสดงดังตารางที่ 3 และตารางที่ 4

ตารางที่ 3 อำนาจการทดสอบของสถิติทดสอบ 5 การทดสอบ ที่ระดับนัยสำคัญ 0.01

	t_4					Gamma(2,0.25)				
n	SF	LF	AD	CVM	PCS	SF	LF	AD	CVM	PCS
20	0.161*	0.075	0.117	0.104	0.042	0.273*	0.129	0.247	0.212	0.131
30	0.230*	0.105	0.165	0.142	0.054	0.465*	0.231	0.434	0.370	0.220
50	0.367*	0.153	0.257	0.232	0.079	0.778*	0.436	0.728	0.643	0.390
70	0.481*	0.213	0.348	0.305	0.107	0.942*	0.641	0.906	0.846	0.630

* หมายถึง อำนาจการทดสอบสูงสุดเมื่อเปรียบเทียบกับสถิติทดสอบอื่นๆ ในกรณีหนึ่งๆ

ตารางที่ 4 อำนาจการทดสอบของสถิติทดสอบ 5 การทดสอบ ที่ระดับนัยสำคัญ 0.05

	t_4					Gamma(2,0.25)				
n	SF	LF	AD	CVM	PCS	SF	LF	AD	CVM	PCS
20	0.290*	0.169	0.234	0.215	0.124	0.488*	0.313	0.451	0.411	0.276
30	0.381*	0.212	0.291	0.267	0.136	0.711*	0.470	0.671	0.607	0.407
50	0.520*	0.298	0.414	0.381	0.200	0.931*	0.695	0.893	0.838	0.655
70	0.632*	0.375	0.509	0.468	0.230	0.987*	0.838	0.971	0.940	0.818

* หมายถึง อำนาจการทดสอบสูงสุดเมื่อเปรียบเทียบกับสถิติทดสอบอื่นๆ ในกรณีหนึ่งๆ

สรุปและอภิปรายผล

จากผลการวิเคราะห์ความสามารถในการควบคุมความคลาดเคลื่อนชนิดที่ 1 และอำนาจการทดสอบของสถิติทดสอบทั้ง 5 การทดสอบ สถิติทดสอบ SF สถิติทดสอบ LF สถิติทดสอบ AD สถิติทดสอบ CVM และ สถิติทดสอบ PCS จากการจำลองข้อมูลตามลักษณะต่างๆ ที่ได้กำหนดไว้ในขอบเขตการวิจัย สรุปผลได้ดังนี้

1. ความสามารถในการควบคุมความคลาดเคลื่อนชนิดที่ 1 ความสามารถในการควบคุมความคลาดเคลื่อนชนิดที่ 1 ของการแจกแจงปรกติที่ N(0,1) และ N(0,4) ที่ระดับนัยสำคัญ 0.01 และ 0.05 สามารถสรุปได้ดังตารางที่ 5

ตารางที่ 5 สถิติทดสอบที่สามารถควบคุมความน่าจะเป็นของความคลาดเคลื่อนชนิดที่ 1 ได้ตามเกณฑ์ของ Bradley [1]

ระดับนัยสำคัญ	n	การแจกแจง	
N(0,1) N(0,4)			
0.01	20	SF,LF,AD,CVM,PCS	SF,LF,AD,CVM,PCS
	30	SF,LF,AD,CVM,PCS	SF,LF,AD,CVM,PCS
	50	SF,LF,AD,CVM,PCS	SF,LF,AD,CVM,PCS
	70	SF,LF,AD,CVM,PCS	SF,LF,AD,CVM,PCS
0.05	20	SF,LF,AD,CVM,PCS	SF,LF,AD,CVM,PCS
	30	SF,LF,AD,CVM,PCS	SF,LF,AD,CVM,PCS
	50	SF,LF,AD,CVM,PCS	SF,LF,AD,CVM,PCS
	70	SF,LF,AD,CVM,PCS	SF,LF,AD,CVM,PCS

SF หมายถึง สถิติทดสอบ Shapiro-Francia

LF หมายถึง สถิติทดสอบ Lilliefor

AD หมายถึง สถิติทดสอบ Anderson-Darling

CVM หมายถึง สถิติทดสอบ Cramer-von Mises

PCS หมายถึง สถิติทดสอบ Pearson chi-square

จากตารางที่ 5 พบว่า สถิติทดสอบ SF LF AD CVM และ PCS สามารถควบคุมความคลาดเคลื่อนชนิดที่ 1 ได้ใกล้เคียงกัน เหมาะสมสำหรับข้อมูลที่มีการแจกแจงปกติที่ N(0,1) และ N(0,4) ที่ระดับนัยสำคัญ 0.01 และ 0.05

2. อำนาจของการทดสอบ

อำนาจการทดสอบของการแจกแจงทีที่ t_4 และการแจกแจงแกมมาที่ Gamma (2 0.25) ที่ระดับนัยสำคัญ 0.01 และ 0.05 สามารถสรุปได้ดังตารางที่ 6

ตารางที่ 6 สถิติทดสอบที่ให้อำนาจการทดสอบสูงที่สุด

ระดับนัยสำคัญ	n	การแจกแจง	
t_4 Gamma (2,0.25)			
0.01	20	SF	SF
	30	SF	SF
	50	SF	SF
	70	SF	SF
0.05	20	SF	SF
	30	SF	SF
	50	SF	SF
	70	SF	SF

SF หมายถึง สถิติทดสอบ Shapiro-Francia

จากตารางที่ 6 พบว่า สถิติทดสอบ SF ให้อำนาจการทดสอบสูงที่สุดเหมาะสมสำหรับข้อมูลที่มีการแจกแจงทีที่ t_4 และการแจกแจงแกมมา Gamma (2,0.25) ที่ระดับนัยสำคัญ 0.01 และ 0.05

ตัวสถิติทดสอบทั้ง 5 ตัวมีประสิทธิภาพในการควบคุมค่าความคลาดเคลื่อนประเภทที่ 1 ได้ในทุกกรณี แต่เมื่อการแจกแจงอื่นที่มีความใกล้เคียงกับการแจกแจงแบบปกติ เช่น การแจกแจงทีมีโอกาสตรวจจับตัวสถิติค่อนข้างยากทำให้เกิดการยอมรับว่าข้อมูลมีการแจกแจงแบบปกติทั้งที่ไม่ได้มีการแจกแจงแบบปกติจึงได้ค่าอำนาจการทดสอบน้อย แต่ถ้ารูปแบบการแจกแจงไม่ได้มีลักษณะสมมาตร เช่น การแจกแจงแกมมา สถิติทดสอบ SF จะสามารถตรวจจับได้ดีที่สุดแสดงว่าข้อมูลไม่ได้มีการแจกแจงแบบปกติทำให้ค่าอำนาจการทดสอบมีค่ามาก

เอกสารอ้างอิง

[1] Bradley, J.V. (1978). Robustness: *Journal of Mathematical and Statistical Psychology.* 31: 144-152.

[2] Cochran, W.G. (1954). Some methods for strengthening the common chi-squared tests. Biometrics.

[3] Royston. (1993). A pocket-calculator algorithm for the Shapiro-Francia W' test for non-Normality: an application to medicine. *Statistics in Medicine.* 12: 181-184.

[4] Shapiro, S.S. and R.S. Francia. (1972). An approxi mate analysis of variance test for normality. J*ournal of the American Statistical Association.* 67: 251-216.

[5] สิริทิพ วะศินรัตน์. (2549). *การเปรียบเทียบการทดสอบของภาวะรูปสนิทดีโดยใช้สถิติไรย์ลิฮูดเรโชว์สำหรับการแจกแจงปกติ.* วิทยานิพนธ์วิทยาศาสตรมหาบัณฑิต (สถิติประยุกต์). กรุงเทพฯ: บัณฑิตวิทยาลัย มหาวิทยาลัยเกษตรศาสตร์.

[6] Stephens, M.A. (1974, September). EDF Statistics for Goodness of fit and Some Comparisons: *Journal of the American Statistical Association.* 69(347): 730-737.

[7] Anderson, T.W. and Darling D.A. (1954). A test of goodness of fit. *Journal of the American Sciences Association.* 49: 765-769.

[8] Pearson, K. (1900). On the criterion that a given system of deviations from probable in the case of a Correlated system of variables is such that it can be reasonably to have arisen from random Sampling. *Phil. Mag.* 50(5): 157-175.