การศึกษาประสิทธิภาพของช่วงความเชื่อมั่นสัมประสิทธิ์การแปรผันควอร์ไทล์

A Study on Efficiency of Confidence Interval for a Coefficient of Quartile Variation

นพรัตน์ แป้นงาม และ อภิศักดิ์ ไชยโรจน์วัฒนา*

Nopparat Panngam and Apisak Chairojwattana*

ภาควิชาคณิตศาสตร์ คณะวิทยาศาสตร์ มหาวิทยาลัยบูรพา

Department of Mathematics, Faculty of Science, Burapha University

Received : 21 September 2017

Accepted : 21 December 2017

Published online : 8 January 2018

บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อศึกษาประสิทธิภาพของช่วงความเชื่อมั่นสัมประสิทธิ์การแปรผันควอร์ไทล์ของ Bonett เมื่อข้อมูลมีการแจกแจงลาปาส การแจกแจงลอจีสติก การแจกแจงไวบูล การแจกแจงล็อกนอร์มอล การแจกแจงปรกติแบบเบ้ และการแจกแจงแกมมา ซึ่งมีลักษณะการแจกแจงเบ้ซ้าย เบ้ขวา และสมมาตรปลายยอดโด่ง ผลการวิจัยพบว่าช่วงความเชื่อมั่นสัมประสิทธิ์การแปรผันควอร์ไทล์ของ Bonett มีประสิทธิภาพดีเมื่อข้อมูลมีการแจกแจงเบ้ขวายกเว้นกรณีที่ขนาดตัวอย่างเท่ากับ 10 และเมื่อข้อมูลมีการแจกแจงในลักษณะเบ้ซ้าย ช่วงความเชื่อมั่นสัมประสิทธิ์การแปรผันควอร์ไทล์ของ Bonett มีประสิทธิภาพดี เฉพาะเมื่อข้อมูลมีขนาดใหญ่และลักษณะการแจกแจงใกล้เคียงสมมาตร

คำสำคัญ : พิสัยควอร์ไทล์ สัมประสิทธิ์ความแปรผัน

Abstract

The objective of this study is to learn on the efficiency of confidence interval for coefficient of quartile variation of Bonett on the data with Laplace Distribution, Logistic Distribution, Weibull Distribution, Lognormal Distribution, Skew Normal Distribution and Gamma Distribution, in event of left-skewed, right-skewed and symmetric leptokurtosis distributions. The study reveals that for right-skewed data, the confidence interval for coefficient of quartile variation of Bonett has good efficiency except for the small sample size (n =10). For left-skewed data, the confidence interval for a coefficient of quartile variation of Bonett has good efficiency and the efficiency increases as sample size increase and the data distribution is closed to symmetry.

Keywords : interquartile range, coefficient of variation

***Corresponding author. E-mail** : apisak@buu.ac.th

บทนำ

การวัดการกระจาย เป็นค่าสถิติ ซึ่งบอกถึงลักษณะการกระจายของข้อมูล การที่ข้อมูลแต่ละชุดมีค่าต่าง ๆ กันนั้นเรียกว่า ข้อมูลมีการกระจาย ถ้าข้อมูลชุดนั้นประกอบด้วยค่าแตกต่างกันมากเรียกว่า ข้อมูลมีการกระจายมาก ถ้าข้อมูลชุดนั้นประกอบด้วยค่าแตกต่างกันน้อย หรือมีค่าใกล้เคียงกันเรียกว่า ข้อมูลมีการกระจายน้อย ถ้าข้อมูลนั้นประกอบด้วยค่าเท่ากันหมดเรียกว่า ข้อมูลไม่มีการกระจาย การวัดการกระจายแบ่งออกเป็น 2 วิธี คือ การวัดการกระจายสัมบูรณ์ (Absolute variation) และการวัดการกระจายสัมพัทธ์ (Relative variation) สำหรับการวัดการกระจายสัมบูรณ์เป็นการวัดการกระจายของข้อมูลเพียงชุดเดียว เพื่อดูว่าข้อมูลชุดนั้นแต่ละค่ามีความแตกต่างกันมากหรือน้อยเพียงใด ได้แก่ พิสัย (Range) ส่วนเบี่ยงเบนควอร์ไทล์ (Quartile deviation) ส่วนเบี่ยงเบนเฉลี่ย (Mean deviation หรือ Average deviation) และส่วนเบี่ยงเบนมาตรฐาน (Standard deviation)

เมื่อต้องการเปรียบเทียบการกระจายของข้อมูลที่มากกว่า 1 ชุด จะใช้วิธีการเปรียบเทียบด้วยการหาค่าการกระจายสัมพัทธ์ โดยใช้อัตราส่วนของค่าที่ได้จากการวัดการกระจายสัมบูรณ์กับค่ากลางของข้อมูลนั้น ๆ เพื่อใช้ในการเปรียบเทียบการกระจายของข้อมูลเหล่านั้น มี 4 วิธี คือ สัมประสิทธิ์พิสัย (Coefficient of range) สัมประสิทธิ์การแปรผันควอร์ไทล์ (Coefficient of quartile variation) สัมประสิทธิ์ส่วนเบี่ยงเบนเฉลี่ย (Coefficient of average deviation) สัมประสิทธิ์การแปรผัน (Coefficient of variation) (Sinsomboonthong, 2004; Niyamangkun, 2005) ซึ่งสัมประสิทธิ์การแปรผัน เป็นค่าสถิติค่าหนึ่งบอกถึงลักษณะการกระจายของข้อมูล หาได้จากอัตราส่วนระหว่างส่วนเบี่ยงเบนมาตรฐานกับค่าเฉลี่ยเลขคณิตของข้อมูลชุดนั้น หากนำสัมประสิทธิ์การแปรผันจากข้อมูลที่ต่างชุดกันมาเปรียบเทียบกัน ข้อมูลชุดใดมีค่าสัมประสิทธิ์การแปรผันสูงกว่าแสดงว่าข้อมูลชุดนั้นมีการกระจายของข้อมูลมากกว่า ข้อมูลชุดใดมีค่าสัมประสิทธิ์การแปรผันต่ำกว่าแสดงว่าข้อมูลชุดนั้นมีการกระจายของข้อมูลน้อยกว่า และถ้าข้อมูลแต่ละชุดมีค่าสัมประสิทธิ์การแปรผันเท่ากันหมดแสดงว่าข้อมูลแต่ละชุดมีการกระจายของข้อมูลเท่ากัน

วิธีการวัดการกระจายที่กล่าวมาข้างต้นเป็นการประมาณค่าแบบจุดทั้งสิ้น สำหรับการประมาณค่าแบบช่วงจากวิธีการวัดการกระจายสัมพัทธ์นั้นได้เริ่มต้นขึ้นโดย McKay (1932) ได้ประมาณการแจกแจงของสัมประสิทธิ์การแปรผันด้วยการแจกแจงที่องศาเสรี n-1 และต่อมาได้มีการประมาณค่าแบบช่วงสำหรับสัมประสิทธิ์การแปรผันโดย Vangel (1996) ซึ่งได้ทำการสร้างช่วงความเชื่อมั่นจากการแจกแจงของสัมประสิทธิ์การแปรผันของ McKay (1932) ช่วงความเชื่อมั่นที่ได้พัฒนามาจาก McKay (1932) และ Vangle (1996) มีความแม่นยำเมื่อตัวอย่างถูกสุ่มมาจากประชากรที่มีการแจกแจงปรกติและสัมประสิทธิ์การแปรผันน้อยกว่า 0.5 (McKay, 1932; Vangel, 1996)

การคำนวณช่วงความเชื่อมั่นสัมประสิทธิ์การแปรผัน โดยทั่วไปเราจะสามารถคำนวณได้จากช่วงความเชื่อมั่นที่ได้พัฒนามาจาก McKay (1932) และ Vangle (1996) แต่ Zar (1984) ได้กล่าวถึงว่าหากข้อมูลมีลักษณะเป็นการแจกแจงแบบเบ้ขวาและมีปลายยอดโด่ง ค่าส่วนเบี่ยงเบนมาตรฐานและค่าเฉลี่ยจะไม่เป็นตัวประมาณค่าที่มีประสิทธิภาพอีกต่อไป โดย Zwillinger and Kokoska (2000) ได้นำเสนอว่าหากตัวอย่างถูกสุ่มมาจากข้อมูลที่มีการแจกแจงที่ไม่ใช่การแจกแจงปรกติ สัมประสิทธิ์การแปรผันควอร์ไทล์จะมีประสิทธิภาพดีกว่าสัมประสิทธิ์การแปรผัน ในขณะที่ Bonett (2006) ได้ทำการหาช่วงความเชื่อมั่นของสัมประสิทธิ์การแปรผันควอร์ไทล์ โดยใช้สัมประสิทธิ์การแปรผันควอร์ไทล์เป็นตัวประมาณค่าและประมาณการแจกแจงด้วยการแจกแจงปรกติ พบว่าผลที่ได้จากช่วงความเชื่อมั่นนี้มีประสิทธิภาพสูงมาก แม้ว่าตัวอย่างจะถูกสุ่มมาจากประชากรที่ไม่มีการแจกแจงปรกติ เมื่อพิจารณาจากความน่าจะเป็นคุ้มรวมใกล้เคียงกับระดับความเชื่อมั่นที่กำหนด อย่างไร

ก็ตามในความเป็นจริงนั้นข้อมูลไม่ได้มีลักษณะการแจกแจงแบบเบ้ขวาเพียงลักษณะเดียว เช่น การแจกแจงของคะแนนสอบสำหรับนักเรียนห้องหนึ่ง ที่นักเรียนส่วนใหญ่ทำคะแนนได้ค่อนข้างมาก แต่มีบางส่วนที่ทำคะแนนได้น้อย ดังนั้นข้อมูลคะแนนสอบของนักเรียนกลุ่มนี้จึงมีการแจกแจงแบบเบ้ซ้าย และจากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องยังไม่พบงานวิจัยที่ศึกษาการวัดการกระจายที่เหมาะสมกับข้อมูลที่มีลักษณะการแจกแจงแบบเบ้ซ้าย การกระจายในลักษณะเบ้ขวา และการกระจายสมมาตรที่มีปลายยอดโด่ง

ผู้วิจัยจึงสนใจที่จะศึกษาประสิทธิภาพของช่วงความเชื่อมั่นสัมประสิทธิ์การแปรผันควอร์ไทล์ของ Bonett (2006) เมื่อข้อมูลมีลักษณะการแจกแจงแบบต่างๆ

วิธีดำเนินการวิจัย

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อศึกษาประสิทธิภาพของช่วงความเชื่อมั่นสัมประสิทธิ์การแปรผันควอร์ไทล์ของ Bonett (2006) เมื่อข้อมูลมีการแจกแจงลาปาส การแจกแจงลอจีสติกส์ การแจกแจงไวบูล การแจกแจงล็อกนอร์มอล การแจกแจงปรกติแบบเบ้ และการแจกแจงแกมมา ซึ่งมีลักษณะการแจกแจงเบ้ซ้าย เบ้ขวา และสมมาตรปลายยอดโด่ง โดย Bonett (2006) เสนอช่วงความเชื่อมั่นสัมประสิทธิ์การแปรผันควอร์ไทล์โดย กำหนดให้ Y เป็นตัวแปรสุ่มต่อเนื่อง ซึ่งมีค่าเป็นบวก ให้ $Y_1, Y_2, \dots, Y_n$ เป็นตัวอย่างสุ่มขนาด n ให้ $\hat{Q}_1$ คือเปอร์เซ็นไทล์ที่ 25 ของตัวอย่าง และ $\hat{Q}_3$ คือเปอร์เซ็นไทล์ที่ 75 ของตัวอย่าง

Stuart and Ord (1994) ได้กำหนดให้

$$var\left(\hat{Q}_1\right) = \frac{3}{16nf_1^2} \qquad (1)$$

$$var\left(\hat{Q}_3\right) = \frac{3}{16nf_3^2} \qquad (2)$$

$$cov\left(\hat{Q}_1, \hat{Q}_3\right) = \frac{1}{16nf_1 f_3} \qquad (3)$$

เมื่อ f_1 เป็นความหนาแน่นความน่าจะเป็นที่ Q_1 และ f_3 เป็นความหนาแน่นความน่าจะเป็นที่ Q_3

Wilks (1959) ได้แสดงไว้ว่าค่าประมาณแบบช่วงที่ 95% ของ Q_1 คือ $\left(Y_{(a)}, Y_{(b)}\right)$ และ ค่าประมาณแบบช่วงที่ 95% ของ Q_3 คือ $\left(Y_{(c)}, Y_{(d)}\right)$

เมื่อ $a = \frac{n}{4} - 1.96\left(\frac{3n}{16}\right)^{\frac{1}{2}}$ $\qquad b = \frac{n}{4} + 1.96\left(\frac{3n}{16}\right)^{\frac{1}{2}}$

$c = n + 1 - b$ $\qquad d = n + 1 - a$

โดยให้ $Y_{(j)}$ เป็นสถิติอันดับที่ j โดยที่ a และ b มีค่าเป็นจำนวนเต็มซึ่ง $a \geq 1$

Hettmansperger and McKean (1998) ได้ประมาณค่าความแปรปรวน $var\left(\hat{Q}_1\right)$ และ $var\left(\hat{Q}_3\right)$ ดังนี้

$$var\left(\hat{Q}_1\right) = \frac{\left(Y_{(b)} - Y_{(a)}\right)^2}{\left(2z_{1-\frac{\alpha^*}{2}}\right)} \qquad (4)$$

$$var\left(\hat{Q}_3\right) = \frac{\left(Y_{(d)} - Y_{(c)}\right)^2}{\left(2z_{1-\frac{\alpha^*}{2}}\right)} \qquad (5)$$

เมื่อ $\alpha^* = 1 - \sum_{i=a}^{b-1} \binom{n}{i} \left(\frac{1}{4}\right)^i \left(\frac{3}{4}\right)^{n-i}$

โดย $z_{1-\frac{\alpha^*}{2}}$ คือ ควอนไทล์ที่ $1-\frac{\alpha^*}{2}$ ของการแจกแจงปรกติมาตรฐาน เมื่อตัวอย่างมีขนาดใหญ่ กำหนดให้ $z_{1-\frac{\alpha^*}{2}} = 1.96$

จาก $var(\hat{Q}_1)$ ของ Stuart and Ord (1994) ในสมการที่ 1 และ $var(\hat{Q}_1)$ ของ Hettmansperger and McKean (1998) ในสมการที่ 4 เราสามารถหาค่าประมาณของ $\hat{f}_1^2$ ได้ โดยกำหนดให้สมการทั้งสองเท่ากัน ดังนี้

$$\frac{3}{n16\,\hat{f}_1^2} = \frac{\left(Y_{(b)} - Y_{(a)}\right)^2}{\left(2z_{1-\frac{\alpha^*}{2}}\right)}$$

$$\hat{f}_1^2 = \frac{3\left(z_{1-\frac{\alpha^*}{2}}\right)^2}{\left\{4n(Y_{(b)} - Y_{(a)})^2\right\}}$$

จาก $var(\hat{Q}_3)$ ของ Stuart and Ord (1994) ในสมการที่ 2 และ $var(\hat{Q}_3)$ ของ Hettmansperger and McKean (1998) ในสมการที่ 5 เราสามารถหาค่าประมาณของ $\hat{f}_3^2$ ได้ โดยกำหนดให้สมการทั้งสองเท่ากัน ดังนี้

$$\frac{3}{n16\,\hat{f}_3^2} = \frac{\left(Y_{(d)} - Y_{(c)}\right)^2}{\left(2z_{1-\frac{\alpha^*}{2}}\right)}$$

$$\hat{f}_3^2 = \frac{3\left(z_{1-\frac{\alpha^*}{2}}\right)^2}{\left\{4n(Y_{(d)} - Y_{(c)})^2\right\}}$$

จาก $cov\left(\hat{Q}_1,\hat{Q}_3\right)$ ของ Stuart and Ord (1994) ในสมการที่ 3 สามารถหาค่าประมาณของ $var\left\{ln\left(\frac{\hat{Q}_3-\hat{Q}_1}{\hat{Q}_3+\hat{Q}_1}\right)\right\}$ ได้ ดังนี้

ให้ $cov\left(\hat{Q}_1,\hat{Q}_3\right) = var(ln(\hat{Q}_3-\hat{Q}_1)) + var(ln(\hat{Q}_3+\hat{Q}_1)) - 2cov(ln(\hat{Q}_3-\hat{Q}_1),ln(\hat{Q}_3+\hat{Q}_1))$

จาก $var(ln(\hat{Q}_3-\hat{Q}_1)) \approx \frac{1}{\left[E(\hat{Q}_3-\hat{Q}_1)\right]^2} var(\hat{Q}_3-\hat{Q}_1)$

$$\approx \frac{1}{(\hat{Q}_3-\hat{Q}_1)^2}\left[var(\hat{Q}_3)+var(\hat{Q}_1)-2cov(\hat{Q}_3,\hat{Q}_1)\right]$$

$$\approx \frac{1}{(\hat{Q}_3-\hat{Q}_1)^2}\left[\frac{3}{n16f_3^2}+\frac{3}{n16f_1^2}-2\left(\frac{1}{n16f_1f_3}\right)\right] \qquad (6)$$

จาก $var(ln(\hat{Q}_3+\hat{Q}_1)) \approx \frac{1}{\left[E(\hat{Q}_3+\hat{Q}_1)\right]^2} var(\hat{Q}_3+\hat{Q}_1)$

$$\approx \frac{1}{(\hat{Q}_3+\hat{Q}_1)^2}\left[var(\hat{Q}_3)+var(\hat{Q}_1)+2cov(\hat{Q}_3,\hat{Q}_1)\right]$$

$$\approx \frac{1}{(\hat{Q}_3+\hat{Q}_1)^2}\left[\frac{3}{n16f_3^2}+\frac{3}{n16f_1^2}+2\left(\frac{1}{n16f_1f_3}\right)\right] \qquad (7)$$

จาก $2cov(ln(\hat{Q}_3-\hat{Q}_1),ln(\hat{Q}_3+\hat{Q}_1)) \approx \frac{2}{\left[E(\hat{Q}_3-\hat{Q}_1)\right]\left[E(\hat{Q}_3+\hat{Q}_1)\right]} var(\hat{Q}_3)+var(\hat{Q}_1)$

$$\approx \frac{2}{(\hat{Q}_3-\hat{Q}_1)(\hat{Q}_3+\hat{Q}_1)}\left[\frac{3}{n16f_3^2}+\frac{3}{n16f_1^2}\right] \qquad (8)$$

จากสมการที่ 6, 7 และ 8 กำหนดให้ $D = \hat{Q}_3-\hat{Q}_1$ และ $S = \hat{Q}_3+\hat{Q}_1$ จะได้ค่าประมาณของ $var\left\{ln\left(\frac{D}{S}\right)\right\}$ แทนด้วย v คือ

$$v = \left(\frac{1}{16n}\right)\left\{\left(\frac{3}{\hat{f}_1^2}+\frac{3}{\hat{f}_3^2}-\frac{2}{\hat{f}_1\hat{f}_3}\right)\Big/D^2+\left(\frac{3}{\hat{f}_1^2}+\frac{3}{\hat{f}_3^2}+\frac{2}{\hat{f}_1\hat{f}_3}\right)\Big/S^2-2\left(\frac{3}{\hat{f}_3^2}-\frac{3}{\hat{f}_1^2}\right)\Big/DS\right\}$$

ดังนั้นช่วงความเชื่อมั่น $(1-\alpha)100\%$ สัมประสิทธิ์การแปรผันควอร์ไทล์คือ

$$exp\left\{ln\left(\frac{D}{S}\right)c \pm z_{1-\frac{\alpha}{2}}v^{\frac{1}{2}}\right\}$$

เมื่อ $c = \frac{n}{n-1}$ เป็นตัวปรับสู่ศูนย์กลางที่ช่วยทำให้ความน่าจะเป็นของความคลาดเคลื่อนที่ปลายหางเท่ากัน

ในงานวิจัยครั้งนี้ได้กำหนดการแจกแจงความน่าจะเป็นและพารามิเตอร์ของการแจกแจงต่าง ๆ เพื่อสร้างข้อมูลที่ใช้ในการทดสอบนั้นมีลักษณะความเบ้และความโด่งแตกต่างกันออกไป โดยจะกำหนดค่าพารามิเตอร์ของการแจกแจงต่าง ๆ ดังต่อไปนี้

1. การแจกแจงลาปาส (Laplace distribution)

พารามิเตอร์ของการแจกแจงนี้ คือ μ เป็นพารามิเตอร์บ่งที่ตั้ง และ b เป็นพารามิเตอร์บ่งขนาด ผู้วิจัยกำหนดค่าพารามิเตอร์ (μ , b) มีค่าเป็น (15, 1) (15, 2) (15, 4) (20, 4) และ (30, 10)

2. การแจกแจงลอจีสติก (Logistic distribution)

พารามิเตอร์ของการแจกแจงนี้ คือ μ เป็นพารามิเตอร์บ่งที่ตั้ง และ s เป็นพารามิเตอร์บ่งขนาด ผู้วิจัยกำหนดค่าพารามิเตอร์ (μ , s) มีค่าเป็น (15, 1) (15, 2) (15, 4) (30, 4) และ (30, 10)

3. การแจกแจงไวบูล (Weibull distribution)

พารามิเตอร์ของการแจกแจงนี้ คือ λ เป็นพารามิเตอร์บ่งขนาด และ k เป็นพารามิเตอร์บ่งรูปร่าง ผู้วิจัยกำหนดค่าพารามิเตอร์ (λ , k) สำหรับการแจกแจงมีลักษณะเบ้ขวา มีค่าเป็น (1, 1) (1, 1.5) (1, 2) (1, 2.5) (1, 3) และ (1, 3.5) สำหรับการแจกแจงมีลักษณะเบ้ซ้าย มีค่าเป็น (1, 7) (1, 20) (1, 200) (1, 5,000) (1, 200,000) และ (1, 2,000,000)

4. การแจกแจงล็อกนอร์มอล (Lognormal distribution)

พารามิเตอร์ของการแจกแจงนี้ คือ μ เป็นพารามิเตอร์บ่งที่ตั้ง และ σ เป็นพารามิเตอร์บ่งขนาด ผู้วิจัยกำหนดค่าพารามิเตอร์ (μ , σ) มีค่าเป็น (0, 0.033) (0, 0.315) (0, 0.552) (0, 5) และ (4, 1.166)

5. การแจกแจงปรกติแบบเบ้ (Skew normal distribution)

พารามิเตอร์ของการแจกแจงนี้ คือ μ เป็นพารามิเตอร์บ่งที่ตั้ง σ เป็นพารามิเตอร์บ่งขนาดและ α เป็นพารามิเตอร์บ่งรูปร่าง ผู้วิจัยกำหนดค่าพารามิเตอร์ (μ , σ , α) สำหรับการแจกแจงมีลักษณะเบ้ขวา มีค่าเป็น (15, 4, 1) (15, 4, 2) (15, 4, 3) (15, 4, 5) (15, 4, 10) และ (15, 4, 10,000) สำหรับการแจกแจงมีลักษณะเบ้ซ้าย มีค่าเป็น (15,1,-1) (15, 1, -2) (15, 1, -3) (15, 1, -5) (15, 1, -10) และ (15, 1, -10,000)

6. การแจกแจงแกมมา (Gamma distribution)

พารามิเตอร์ของการแจกแจงนี้ คือ θ เป็นพารามิเตอร์บ่งขนาด และ k เป็นพารามิเตอร์บ่งรูปร่าง ผู้วิจัยกำหนดค่าพารามิเตอร์ (θ , k) มีค่าเป็น (1, 0.4) (1, 1) (1, 4) (1, 16) (2, 4)

ขั้นตอนการดำเนินการวิจัยดังนี้

1. กำหนดให้ตัวแปรสุ่ม X เป็นตัวแปรสุ่มที่มีการแจกแจงที่กำหนดในแผนการดำเนินงานวิจัย

2. คำนวณค่าเฉลี่ย ความแปรปรวน และสัมประสิทธิ์ความเบ้ ของตัวแปรสุ่ม X จากชุดของพารามิเตอร์ที่กำหนดไว้

3. คำนวณหาค่าสัมประสิทธิ์การแปรผันควอร์ไทล์ (Coefficient of quartile variation : cqv) ในแต่ละชุดของพารามิเตอร์ที่กำหนดไว้โดยค่า cqv สามารถหาได้จาก

$$cqv = \frac{(Q_3 - Q_1)}{(Q_3 + Q_1)}$$

โดยที่ Q_1 เป็นค่าเปอร์เซ็นไทล์ที่ 25 ของการแจกแจงที่สนใจศึกษา

Q_3 เป็นค่าเปอร์เซ็นไทล์ที่ 75 ของการแจกแจงที่สนใจศึกษา

4. จำลองข้อมูลของตัวแปรสุ่ม X ตามขนาดตัวอย่าง โดยกำหนดขนาดตัวอย่าง มีค่าเท่ากับ 10 25 50 และ 100 และ ค่าสัมประสิทธิ์ความเชื่อมั่นให้มีค่าเท่ากับ 0.90 0.95 และ 0.99

5. คำนวณช่วงความเชื่อมั่นสัมประสิทธิ์การแปรผันควอร์ไทล์ด้วยช่วงความเชื่อมั่นของ Bonett (2006)

6. ทำซ้ำขั้นตอนที่ 4 และ 5 จำนวน 50,000 ครั้ง

7. คำนวณค่าความน่าจะเป็นคุ้มรวม (Coverage probability: CP) ซึ่งหาได้จาก

$$CP = \frac{\text{จำนวนครั้งทั้งหมดที่ช่วงความเชื่อมั่นครอบคลุมค่าพารามิเตอร์}}{\text{จำนวนรอบที่ศึกษา}}$$

8. ทำซ้ำขั้นตอนที่ 1 ถึง 7 จนครบทุกสถานการณ์

เกณฑ์ที่ใช้ในการเปรียบเทียบ

ในการเปรียบเทียบประสิทธิภาพของช่วงความเชื่อมั่นสัมประสิทธิ์การแปรผันควอร์ไทล์นั้น จะพิจารณาจากความน่าจะเป็นคุ้มรวม โดยมีเกณฑ์ในการเปรียบเทียบคือ ช่วงความเชื่อมั่นที่มีประสิทธิภาพดี จะมีค่าความน่าจะเป็นคุ้มรวมใกล้เคียงกับสัมประสิทธิ์ความเชื่อมั่นที่กำหนด

ผลการวิจัยและวิจารณ์ผล

เมื่อผู้วิจัยได้จำลองข้อมูลจำนวน 50,000 ชุดในแต่ละการแจกแจง ค่าพารามิเตอร์ และขนาดตัวอย่างตามที่กำหนดแล้ว และคำนวณค่า cqv และช่วงความเชื่อมั่นสัมประสิทธิ์การแปรผันส่วนเบี่ยงเบนควอร์ไทล์ของ Bonett (2006) จากนั้นนำผลที่ได้ไปคำนวณหาค่าความน่าจะเป็นคุ้มรวม ซึ่งสามารถสรุปแยกตามการแจกแจงได้ดังแสดงในตารางที่ 1 – 8

ตารางที่ 1 ค่าความน่าจะเป็นคุ้มรวมของการแจกแจงลาปาส

พารามิเตอร์ (μ , b)	cqv	ความน่าจะเป็นคุ้มรวม (CP)			
		n = 10	n = 25	n = 50	n = 100
		เมื่อสัมประสิทธิ์ความเชื่อมั่นเท่ากับ 0.90			
15, 1	0.046	0.905	0.938	0.911	0.924
15, 2	0.092	0.916	0.940	0.912	0.923
15, 4	0.185	0.916	0.937	0.911	0.924
20, 4	0.092	0.918	0.940	0.912	0.923
30, 10	0.231	0.916	0.938	0.907	0.923
		เมื่อสัมประสิทธิ์ความเชื่อมั่นเท่ากับ 0.95			
15, 1	0.046	0.939	0.957	0.964	0.958
15, 2	0.092	0.947	0.959	0.963	0.957
15, 4	0.185	0.946	0.956	0.962	0.958
20, 4	0.139	0.948	0.957	0.963	0.956
30, 10	0.231	0.944	0.957	0.961	0.950
		เมื่อสัมประสิทธิ์ความเชื่อมั่นเท่ากับ 0.99			
15, 1	0.046	0.963	0.999	0.996	0.995
15, 2	0.092	0.966	0.998	0.995	0.995
15, 4	0.185	0.967	0.998	0.995	0.994
20, 4	0.092	0.966	0.998	0.995	0.995
30, 10	0.231	0.965	0.998	0.994	0.994

หมายเหตุ : การแจกแจงลาปาสมีค่าความเบ้=0 และความโด่ง=3 ทุกกรณี

ตารางที่ 2 ค่าความน่าจะเป็นคุ้มรวมของการแจกแจงลอจีสติก

พารามิเตอร์ (μ , s)	cqv	**ความน่าจะเป็นคุ้มรวม (CP)**			
		n = 10	n = 25	n = 50	n = 100
		เมื่อสัมประสิทธิ์ความเชื่อมั่นเท่ากับ 0.90			
15, 1	0.073	0.892	0.932	0.908	0.923
15, 2	0.146	0.909	0.933	0.908	0.922
15, 4	0.293	0.907	0.926	0.902	0.919
30, 4	0.146	0.907	0.936	0.907	0.922
30, 10	0.366	0.903	0.923	0.901	0.915
		เมื่อสัมประสิทธิ์ความเชื่อมั่นเท่ากับ 0.95			
15, 1	0.073	0.937	0.956	0.962	0.956
15, 2	0.146	0.937	0.955	0.960	0.955
15, 4	0.293	0.949	0.954	0.959	0.951
30, 4	0.146	0.946	0.957	0.961	0.955
30, 10	0.366	0.946	0.951	0.955	0.950
		เมื่อสัมประสิทธิ์ความเชื่อมั่นเท่ากับ 0.99			
15, 1	0.073	0.964	0.999	0.996	0.995
15, 2	0.146	0.971	0.999	0.996	0.994
15, 4	0.293	0.974	0.998	0.994	0.994
30, 4	0.146	0.972	0.999	0.996	0.994
30, 10	0.366	0.976	0.998	0.995	0.993

หมายเหตุ : การแจกแจงลอจีสติกมีค่าความเบ้=0 และความโด่ง=1.2 ทุกกรณี

ตารางที่ 3 ค่าความน่าจะเป็นคุ้มรวมของการแจกแจงไวบูล (เบ้ขวา)

พารามิเตอร์ (λ , k)	ความเบ้ , ความโด่ง	cqv	ความน่าจะเป็นคุ้มรวม (CP)			
			n = 10	n = 25	n = 50	n = 100
			เมื่อสัมประสิทธิ์ความเชื่อมั่นเท่ากับ 0.90			
1, 1	2, 6	0.656	0.898	0.916	0.899	0.917
1, 1.5	1.073, 1.390	0.481	0.899	0.921	0.902	0.920
1, 2	0.631, 0.245	0.374	0.899	0.922	0.903	0.920
1, 2.5	0.359, -0.143	0.305	0.898	0.923	0.903	0.916
1, 3	0.168, -0.271	0.256	0.895	0.923	0.899	0.918
1, 3.5	0.025, -0.287	0.221	0.894	0.922	0.901	0.918
			เมื่อสัมประสิทธิ์ความเชื่อมั่นเท่ากับ 0.95			
1, 1	2, 6	0.656	0.958	0.951	0.953	0.949
1, 1.5	1.073, 1.390	0.481	0.953	0.953	0.956	0.952
1, 2	0.631, 0.245	0.374	0.953	0.953	0.956	0.952
1, 2.5	0.359, -0.143	0.305	0.950	0.950	0.956	0.952
1, 3	0.168, -0.271	0.256	0.946	0.951	0.955	0.952
1, 3.5	0.025, -0.287	0.221	0.943	0.950	0.955	0.953
			เมื่อสัมประสิทธิ์ความเชื่อมั่นเท่ากับ 0.99			
1, 1	2, 6	0.656	0.989	0.991	0.990	0.991
1, 1.5	1.073, 1.390	0.481	0.984	0.994	0.992	0.992
1, 2	0.631, 0.245	0.374	0.980	0.996	0.993	0.992
1, 2.5	0.359, -0.143	0.305	0.976	0.997	0.994	0.993
1, 3	0.168, -0.271	0.256	0.974	0.997	0.994	0.993
1, 3.5	0.025, -0.287	0.221	0.971	0.997	0.995	0.993

ตารางที่ 4 **ค่าความน่าจะเป็นคุ้มรวม (CP) ของการแจกแจงไวบูล (เบ้ซ้าย)**

พารามิเตอร์ (λ, k)	ความเบ้ , ความโด่ง	cqv	ความน่าจะเป็นคุ้มรวม (CP) n = 10	n = 25	n = 50	n = 100
			เมื่อสัมประสิทธิ์ความเชื่อมั่นเท่ากับ 0.90			
1, 2,000,000	-1682.661, 4.6×10^{9}	3.9×10^{-7}	0.036	0.268	0.508	0.723
1, 200,000	-1.263, 1.3×10^{5}	3.9×10^{-6}	0.079	0.419	0.613	0.781
1, 5,000	-1.138, 2.494	1.5×10^{-4}	0.280	0.674	0.762	0.857
1, 200	-1.110, 2.259	0.004	0.618	0.842	0.853	0.895
1, 20	-0.868, 1.267	0.039	0.827	0.905	0.889	0.913
1, 7	-0.463, 0.187	0.112	0.872	0.916	0.897	0.919
			เมื่อสัมประสิทธิ์ความเชื่อมั่นเท่ากับ 0.95			
1, 2,000,000	-1682.661, 4.6×10^{9}	3.9×10^{-7}	0.047	0.327	0.667	0.811
1, 200,000	-1.263, 1.3×10^{5}	3.9×10^{-6}	0.102	0.491	0.765	0.857
1, 5,000	-1.138, 2.494	1.5×10^{-4}	0.363	0.741	0.878	0.912
1, 200	-1.110, 2.259	0.004	0.722	0.882	0.932	0.939
1, 20	-0.868, 1.267	0.039	0.892	0.938	0.954	0.950
1, 7	-0.463, 0.187	0.112	0.931	0.948	0.957	0.951
			เมื่อสัมประสิทธิ์ความเชื่อมั่นเท่ากับ 0.99			
1, 2,000,000	-1682.661, 4.6×10^{9}	3.9×10^{-7}	0.058	0.911	0.926	0.959
1, 200,000	-1.263, 1.3×10^{5}	3.9×10^{-6}	0.137	0.961	0.957	0.974
1, 5,000	-1.138, 2.494	1.5×10^{-4}	0.461	0.992	0.985	0.987
1, 200	-1.110, 2.259	0.004	0.809	0.998	0.995	0.992
1, 20	-0.868, 1.267	0.039	0.934	0.999	0.995	0.994
1, 7	-0.463, 0.187	0.112	0.960	0.998	0.997	0.994

ตารางที่ 5 ค่าความน่าจะเป็นคุ้มรวม (CP) ของการแจกแจงล็อกนอร์มอล

พารามิเตอร์ (μ , σ)	**ความเบ้ , ความโด่ง**	cqv	**ความน่าจะเป็นคุ้มรวม (CP)**			
			n = 10	n = 25	n = 50	n = 100
			เมื่อสัมประสิทธิ์ความเชื่อมั่นเท่ากับ 0.90			
0, 5	$1.9{\times}10^{16}$,$-2.7{\times}10^{43}$	0.998	0.936	0.925	0.916	0.922
0, 1.166	10.029,387.667	0.656	0.914	0.931	0.914	0.927
4, 1.166	10.029, 387.667	0.656	0.916	0.932	0.917	0.926
0, 0.552	2.003,7.890	0.356	0.917	0.935	0.910	0.925
0, 0.315	1.003,1.839	0.209	0.907	0.934	0.909	0.925
0, 0.033	0.099,0.017	0.022	0.787	0.897	0.888	0.913
			เมื่อสัมประสิทธิ์ความเชื่อมั่นเท่ากับ 0.95			
0, 5	$1.9{\times}10^{16}$, $-2.7{\times}10^{43}$	0.998	0.975	0.952	0.947	0.943
0, 1.166	10.029, 387.667	0.656	0.958	0.954	0.958	0.954
4, 1.166	10.029, - 387.667	0.656	0.958	0.954	0.959	0.958
0, 0.552	2.003, 7.890	0.356	0.957	0.957	0.960	0.956
0, 0.315	1.003, 1.839	0.209	0.953	0.958	0.961	0.957
0, 0.033	0.099, 0.017	0.022	0.872	0.935	0.952	0.952
			เมื่อสัมประสิทธิ์ความเชื่อมั่นเท่ากับ 0.99			
0, 5	$1.9{\times}10^{16}$, $-2.7{\times}10^{43}$	0.998	0.995	0.993	0.986	0.981
0, 1.166	10.029, 387.667	0.656	0.986	0.994	0.990	0.991
4, 1.166	10.029, -387.667	0.656	0.987	0.993	0.990	0.992
0, 0.552	2.003, 7.890	0.356	0.981	0.997	0.993	0.994
0, 0.315	1.003, 1.839	0.209	0.978	0.998	0.995	0.994
0, 0.033	0.099, 0.017	0.022	0.926	0.999	0.996	0.995

ตารางที่ 6 **ค่าความน่าจะเป็นคุ้มรวม (CP) ของการแจกแจงปรกติแบบเบ้ (เบ้ขวา)**

พารามิเตอร์ (μ , σ , α)	ความเบ้ , ความโด่ง	cqv	ความน่าจะเป็นคุ้มรวม (CP)			
			n = 10	n = 25	n = 50	n = 100
			เมื่อสัมประสิทธิ์ความเชื่อมั่นเท่ากับ 0.90			
15, 4,100,000	0.995, 0.869	0.093	0.835	0.908	0.896	0.914
15, 4, 10	0.956, 0.823	0.093	0.846	0.912	0.896	0.914
15, 4, 5	0.851, 0.705	0.093	0.865	0.918	0.898	0.915
15, 4, 3	0.667, 0.510	0.097	0.881	0.923	0.905	0.922
15, 4, 2	0.454, 0.305	0.105	0.892	0.930	0.906	0.921
15, 4, 1	0.137, 0.062	0.129	0.894	0.931	0.905	0.922
			เมื่อสัมประสิทธิ์ความเชื่อมั่นเท่ากับ 0.95			
15, 4,100,000	0.995, 0.869	0.093	0.905	0.941	0.954	0.952
15, 4, 10	0.956, 0.823	0.093	0.910	0.942	0.953	0.950
15, 4, 5	0.851, 0.705	0.093	0.923	0.947	0.956	0.952
15, 4, 3	0.667, 0.510	0.097	0.936	0.952	0.959	0.957
15, 4, 2	0.454, 0.305	0.105	0.939	0.956	0.961	0.955
15, 4, 1	0.137, 0.062	0.129	0.944	0.955	0.961	0.955
			เมื่อสัมประสิทธิ์ความเชื่อมั่นเท่ากับ 0.99			
15, 4,100,000	0.995, 0.869	0.093	0.945	0.998	0.995	0.993
15, 4, 10	0.956, 0.823	0.093	0.950	0.998	0.995	0.993
15, 4, 5	0.851, 0.705	0.093	0.959	0.999	0.995	0.994
15, 4, 3	0.667, 0.510	0.097	0.964	0.998	0.995	0.994
15, 4, 2	0.454, 0.305	0.105	0.967	0.999	0.995	0.994
15, 4, 1	0.137, 0.062	0.129	0.970	0.999	0.995	0.994

ตารางที่ 7 ค่าความน่าจะเป็นคุ้มรวม (CP) ของการแจกแจงปรกติแบบเบ้ (เบ้ซ้าย)

พารามิเตอร์ (μ , σ , α)	ความเบ้ , ความโด่ง	cqv	ความน่าจะเป็นคุ้มรวม (CP)			
			n = 10	n = 25	n = 50	n = 100
			เมื่อสัมประสิทธิ์ความเชื่อมั่นเท่ากับ 0.90			
15, 4, -100,000	-0.995, 0.869	0.029	0.757	0.880	0.876	0.905
15, 4, -10	-0.956, 0.823	0.029	0.766	0.882	0.876	0.906
15, 4, -5	-0.851, 0.705	0.029	0.780	0.890	0.882	0.908
15, 4, -3	-0.667, 0.510	0.030	0.798	0.901	0.884	0.909
15, 4, -2	-0.454, 0.305	0.032	0.820	0.908	0.892	0.914
15, 4, -1	-0.137, 0.062	0.038	0.833	0.915	0.893	0.917
			เมื่อสัมประสิทธิ์ความเชื่อมั่นเท่ากับ 0.95			
15, 4,-100,000	-0.995, 0.869	0.029	0.836	0.915	0.942	0.944
15, 4, -10	-0.956, 0.823	0.029	0.845	0.916	0.943	0.945
15, 4, -5	-0.851, 0.705	0.029	0.857	0.923	0.947	0.945
15, 4, -3	-0.667, 0.510	0.030	0.876	0.931	0.950	0.948
15, 4, -2	-0.454, 0.305	0.032	0.888	0.938	0.955	0.951
15, 4, -1	-0.137, 0.062	0.038	0.902	0.941	0.957	0.953
			เมื่อสัมประสิทธิ์ความเชื่อมั่นเท่ากับ 0.99			
15, 4,-100,000	-0.995, 0.869	0.029	0.895	0.997	0.993	0.993
15, 4, -10	-0.956, 0.823	0.029	0.898	0.998	0.994	0.992
15, 4, -5	-0.851, 0.705	0.029	0.912	0.998	0.994	0.992
15, 4, -3	-0.667, 0.510	0.030	0.926	0.999	0.995	0.994
15, 4, -2	-0.454, 0.305	0.032	0.936	0.999	0.996	0.995
15, 4, -1	-0.137, 0.062	0.038	0.944	0.999	0.996	0.995

ตารางที่ 8 **ค่าความน่าจะเป็นคุ้มรวมของการแจกแจงแกมมา**

พารามิเตอร์ (θ, k)	**ความเบ้ , ความโด่ง**	cqv	**ความน่าจะเป็นคุ้มรวม (CP)**			
			n = 10	n = 25	n = 50	n = 100
			เมื่อสัมประสิทธิ์ความเชื่อมั่นเท่ากับ 0.90			
1, 0.4	3.162, 15	0.911	0.898	0.902	0.893	0.913
1, 1	1, 1.5	0.656	0.897	0.917	0.902	0.918
1, 4	2, 6	0.337	0.910	0.932	0.910	0.922
1, 16	0.5, 0.375	0.169	0.901	0.930	0.905	0.921
1, 400	0.1, 0.015	0.034	0.822	0.910	0.892	0.918
2, 4	1, 1.5	0.337	0.909	0.932	0.907	0.922
			เมื่อสัมประสิทธิ์ความเชื่อมั่นเท่ากับ 0.95			
1, 0.4	3.162, 15	0.911	0.970	0.948	0.947	0.941
1, 1	1, 1.5	0.656	0.959	0.950	0.952	0.949
1, 4	2, 6	0.337	0.955	0.957	0.961	0.955
1, 16	0.5, 0.375	0.169	0.950	0.957	0.961	0.956
1, 400	0.1, 0.015	0.034	0.895	0.940	0.956	0.954
2, 4	1, 1.5	0.337	0.958	0.957	0.959	0.956
			เมื่อสัมประสิทธิ์ความเชื่อมั่นเท่ากับ 0.99			
1, 0.4	3.162, 15	0.911	0.995	0.986	0.984	0.985
1, 1	1, 1.5	0.656	0.989	0.990	0.990	0.990
1, 4	2, 6	0.337	0.983	0.997	0.994	0.994
1, 16	0.5, 0.375	0.169	0.975	0.998	0.996	0.994
1, 400	0.1, 0.015	0.034	0.940	0.999	0.996	0.995
2, 4	1, 1.5	0.337	0.982	0.997	0.994	0.993

เมื่อพิจารณาการแจกแจงลาปาสและการแจกแจงลอจีสติก พบว่า ผลการทดลองที่ได้มีลักษณะคล้ายคลึงกัน โดยที่สัมประสิทธิ์ความเชื่อมั่น 0.90 ค่าความน่าจะเป็นคุ้มรวมจะมีค่าใกล้เคียงกับค่าสัมประสิทธิ์ความเชื่อมั่น เมื่อขนาดตัวอย่างเท่ากับ 10, 50 และ 100 ที่สัมประสิทธิ์ความเชื่อมั่น 0.95 ค่าความน่าจะเป็นคุ้มรวมจะมีค่าใกล้เคียงกับค่าสัมประสิทธิ์ความเชื่อมั่นในทุกกรณี ในขณะที่สัมประสิทธิ์ความเชื่อมั่น 0.99 นั้น ค่าความน่าจะเป็นคุ้มรวมจะมีค่าใกล้เคียงกับค่าสัมประสิทธิ์ความเชื่อมั่น เมื่อขนาดตัวอย่างเท่ากับ 25, 50 และ 100

สำหรับการแจกแจงไวบูล (เบ้ขวา) การแจกแจงล็อกนอร์มอล การแจกแจงปรกติแบบเบ้ (เบ้ขวา) และการแจกแจงแกมมา พบว่า ค่าความน่าจะเป็นคุ้มรวมมีค่าใกล้เคียงกับค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด เมื่อขนาดตัวอย่างเท่ากับ 25, 50 และ 100 และมีค่าความเบ้เพิ่มมากขึ้น

สำหรับการแจกแจงไวบูล (เบ้ซ้าย) และการแจกแจงปรกติแบบเบ้ (เบ้ซ้าย) พบว่า ค่าความน่าจะเป็นคุ้มรวมมีค่าใกล้เคียงกับค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด เมื่อขนาดตัวอย่างเพิ่มมากขึ้นและลักษณะการแจกแจงใกล้เคียงสมมาตร

สรุปผลการวิจัย

จากผลการวิจัยสามารถสรุปได้ว่า ช่วงความเชื่อมั่นสัมประสิทธิ์การแปรผันควอร์ไทล์ของ Bonett (2006) มีประสิทธิภาพดีเมื่อข้อมูลมีการแจกแจงเบ้ขวายกเว้นกรณีที่ขนาดตัวอย่างเท่ากับ 10 ซึ่งสอดคล้องกับงานวิจัยของ Bonett (2006) แต่อย่างไรก็ตามเมื่อข้อมูลมีการแจกแจงในลักษณะเบ้ซ้าย ซึ่ง Bonett (2006) ไม่ได้ทำการศึกษา พบว่าช่วงความเชื่อมั่นสัมประสิทธิ์การแปรผันควอร์ไทล์ของ Bonett (2006) มีประสิทธิภาพดีเมื่อขนาดตัวอย่างใหญ่และลักษณะการแจกแจงใกล้เคียงสมมาตร

กิตติกรรมประกาศ

ขอขอบคุณคณะวิทยาศาสตร์ มหาวิทยาลัยบูรพา ที่ให้การสนับสนุนทุนสำหรับการวิจัยครั้งนี้ และผู้วิจัยขอขอบคุณ ดร.จุฑาพร เนียมวงษ์ ที่ให้คำปรึกษา

เอกสารอ้างอิง

Bonett, D.G., (2006). Confidence interval for a coefficient of quartile variation. Computational Statistics & Data Analysis, *50*, 2953 – 2957.

Niyamangkun, S. (2005). Statistics for Research. (2nd Edition). Bangkok : Kasetsart University Press. (in Thai)

McKay, A.T., (1932). Distribution of the coefficient of variation and extended ‘t’ distribution. Journal of the Royal Statistical Society, *95*, 695 – 698.

Sinsomboonthong, S. (2004). Elementary statistics, (5th Edition). Bangkok : Chamchuri Products. (in Thai)

Vangel, M.G. (1996). Confidence intervals for a normal coefficient of variation. The American Statistician, *50*, 21 – 26.

Wilks, S.S. (1959). Nonparametric Statistical Inference. Wiley, New York.

Zar, J.H. (1984). Biostatistical Analysis. second ed. Prentice-Hall, Englewood Cliffs, NJ.

Hettmansperger, T.P., McKean, J.W., (1998). Robust Nonparametric Statistical Methods. Wiley, London.

Stuart, A., Ord, J.K., (1994). Kendalls’ Advanced Theory of Statistics, Volume I: Distribution Theory. sixth ed. Edward Amold, London

Zwillinger, D., Kokoska, S. (2000). Standard Probability and Statistical Tables and Formula. Chapman & Hall, Boca Raton.