การปรับปรุงช่วงความเชื่อมั่นสำหรับการประมาณค่าพารามิเตอร์ของการแจกแจงปัวซง
Improving the Confidence Intervals for Parameter Estimation of Poisson Distribution.

ศุภกานต์ ศรีวิชัย* และ มานัดถุ์ คำกอง
ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยเชียงใหม่
Supakan Sriwichai* and Manad Khamkong
Department of Statistics, Faculty of Science, Chiangmai University.

บทคัดย่อ

การศึกษาครั้งนี้มีวัตถุประสงค์เพื่อนำเสนอวิธีการปรับปรุงช่วงความเชื่อมั่นสำหรับการประมาณค่าพารามิเตอร์ของการแจกแจงปัวซงโดยศึกษาวิธีการประมาณค่าช่วงความเชื่อมั่นจาก 3 วิธี คือ วิธีประมาณแบบ Score (Score) วิธีของ Wald แบบปรับค่าความต่อเนื่อง (Wald Interval with Continuity Correction) วิธีการแปลงข้อมูลที่ทำให้ความแปรปรวนคงที่ (Variance Stabilizing) และผู้วิจัยได้เสนอวิธีการปรับปรุงช่วงความเชื่อมั่นโดยประมาณด้วยการแจกแจงที (t-distribution) ของทั้ง 3 วิธีนี้ การจำลองข้อมูลเพื่อทำการเปรียบเทียบค่าความน่าจะเป็นครอบคลุมและค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่นของแต่ละสถานการณ์ เมื่อกำหนดขนาดตัวอย่างเท่ากับ 10, 15, 30, 50 และ 100 ที่ค่าพารามิเตอร์ เท่ากับ 1, 1.5, 3 และ 5 ที่ระดับช่วงความเชื่อมั่น 95% โดยทำการทดลองซ้ำ 10,000 ครั้งในแต่ละสถานการณ์ ผลการวิจัยสรุปได้ว่า วิธีการปรับปรุงช่วงความเชื่อมั่นโดยวิธีประมาณแบบ Score ด้วยการแจกแจงที ให้ค่าความน่าจะเป็นครอบคลุมระดับความเชื่อมั่นที่กำหนดสูงสุดและให้ค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่นต่ำสุดในทุกสถานการณ์

คำสำคัญ : ช่วงความเชื่อมั่น การแจกแจงปัวซง ค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่น ค่าความน่าจะเป็นครอบคลุม

Abstract

The purpose of this study is to propose an improving the confidence intervals for parameter estimation of the Poisson distribution based on the asymptotic t-distribution by studying how to construct the confidence intervals from the Score method, the Wald with continuity correction method and the variances stabilizing methods, then compare the coverage probabilities estimation and the average widths estimation between the 3 confidence intervals. The comparisons were done using 10,000 random samples from the asymptotic t-distribution with sample sizes (n) of 10, 15, 30, 50 and 100 while values of parameter (λ) of 1, 1.5, 3 and 5 are chosen, all of which are considered at 95% confidence interval. It is shown that confidence intervals generated with the Score method have coverage probabilities close to 95% confidence interval and have shortest average widths in all situations.

Keywords : confidence interval, Poisson Distribution, Coverage Probability, Expected Width of Confidence Interval

***Corresponding author.** E-mail: krootannaraak@hotmail.com

บทนำ

ปัจจุบันมีงานวิจัยหลายแขนงสนใจศึกษาหรือทำการทดลองเกี่ยวกับสถานการณ์ต่างๆ ที่เกิดขึ้นในช่วงเวลาใดเวลาหนึ่งหรือขอบเขตใดขอบเขตหนึ่งโดยสนใจจำนวนของเหตุการณ์ที่จะเกิดขึ้น และได้นำความรู้เกี่ยวกับการแจกแจงปัวซงไปประยุกต์ใช้กับงานต่างๆ อย่างแพร่หลายทั้งในด้านที่เกี่ยวข้องกับการแพทย์ การเกษตร วิศวกรรมศาสตร์ ด้านบริหารธุรกิจ รวมถึงด้านอื่นๆ ที่เกี่ยวข้องกับข้อมูลที่เป็นจำนวนนับภายใต้กรอบเวลาที่กำหนดหรือขอบเขตที่สนใจ เช่น จำนวนผู้ป่วยที่เข้ารับการรักษาในโรงพยาบาลแห่งหนึ่งในระยะเวลาหนึ่งเดือน จำนวนรถยนต์ที่สูญหายในมหาวิทยาลัยเชียงใหม่ในปี พ.ศ. 2553 จำนวนเม็ดเลือดขาวในเลือด 1 ลบ.ซม. จำนวนลูกค้าที่มาใช้บริการในช่องทางชำระเงินในเวลา 1 ชั่วโมง เป็นต้น ในการจัดทำข้อสรุปเกี่ยวกับประชากรจะต้องอาศัยการเก็บรวบรวมข้อมูลจากประชากรที่ศึกษาแล้วนำข้อมูลเหล่านั้นมาทำการประมวลผลเพื่อนำผลไปอธิบายประชากรที่กำลังศึกษา ในความเป็นจริงการเก็บรวบรวมข้อมูลจากทุกหน่วยประชากรเป็นไปได้ยากเนื่องจากมีข้อจำกัดเกี่ยวกับเวลาและค่าใช้จ่าย จึงหาข้อสรุปของประชากรจากการสุ่มตัวอย่างเพื่อนำมาประมาณค่าพารามิเตอร์ที่สนใจเพื่อใช้ในการหาข้อสรุปต่อไป ซึ่งในระเบียบวิธีทางสถิติที่สำคัญที่ใช้ในการจัดทำข้อสรุปสำหรับพารามิเตอร์ที่สนใจ มี 2 วิธี คือการประมาณค่าแบบจุด (Point Estimation) และการประมาณค่าแบบช่วง (Interval Estimation)

ในการสร้างตัวประมาณแบบช่วงสำหรับพารามิเตอร์ λ ของการแจกแจงปัวซงที่ผ่านมานั้น ส่วนใหญ่อาศัยการสร้างตัวประมาณแบบช่วงภายใต้ตัวอย่างขนาดใหญ่ (n มีค่ามาก) ซึ่งอาศัยทฤษฎีบทขีดจำกัดสู่ส่วนกลาง (Central Limit Theorem : CLT) มาช่วยในการสร้างช่วงความเชื่อมั่น [Lehmann (2005)] แต่เมื่อจำนวนตัวอย่างสุ่มมีขนาดเล็ก เช่น ในทางการแพทย์ที่สนใจศึกษาโรคที่พบได้ยาก แต่ประสบปัญหาคือไม่สามารถหาตัวอย่างสุ่มคนไข้ที่ป่วยเป็นโรคพบยากชนิดนี้ได้จำนวนมากเพียงพอ จึงเป็นอุปสรรคในการหาข้อสรุปของสมมติฐานที่อาจจะมีความคลาดเคลื่อนในการหาข้อสรุปได้ แต่นักวิจัยส่วนใหญ่ที่ประสบปัญหาขนาดตัวอย่างเล็กก็ยังคงใช้ทฤษฎีบทขีดจำกัดสู่ส่วนกลางในการสร้างตัวประมาณแบบช่วงสำหรับประมาณค่าเฉลี่ยของตัวอย่างสุ่มดังกล่าว ดังนั้น ช่วงความเชื่อมั่นที่ได้จึงไม่สามารถให้ค่าความน่าจะเป็นของการครอบคลุม (Coverage Probability) เท่ากับ 1-α ได้พอดี นั่นคือไม่สามารถให้ค่าประมาณที่ถูกต้องและสอดคล้องกับข้อตกลงของทฤษฎี เนื่องจากตัวอย่างมีขนาดเล็ก

ดังนั้น ผู้วิจัยจึงเสนอทางเลือกในการปรับปรุงช่วงความเชื่อมั่นสำหรับการประมาณค่าพารามิเตอร์ λ ของประชากรที่มีการแจกแจงปัวซง โดยประยุกต์ใช้หลักการของ Pan (2002) ในกรณีที่ขนาดตัวอย่างเล็กและพารามิเตอร์มีค่าน้อยเพื่อเป็นแนวทางในการเลือกใช้วิธีการประมาณช่วงความเชื่อมั่นสำหรับพารามิเตอร์ λ ที่เหมาะสมสำหรับแต่ละสถานการณ์ของค่า n และ λ

ขั้นตอนของการศึกษา

ในการวิจัยครั้งนี้มีวิธีดำเนินการวิจัยผ่านการศึกษาแนวทางในการปรับปรุงช่วงความเชื่อมั่นโดยประยุกต์ใช้วิธีการของ Satterthwaite's method จากการศึกษาจากงานวิจัยของ Pan (2002) ที่ได้ศึกษาการประมาณช่วงความเชื่อมั่นของการแจกแจงทวินาม โดยประมาณเข้าสู่การแจกแจงทีโดยทำการปรับค่าองศาเสรีเพื่อให้ได้ค่าประมาณช่วงความเชื่อมั่นที่แคบที่สุดผู้วิจัยจึงได้ทำการศึกษาโดยจะทำประยุกต์งานวิจัยของ Pan (2002) มาใช้ในการปรับค่าองศาเสรีของการแจกแจงปัวซง มีกระบวนการดังนี้

$$c = \frac{Var(Var(\hat{\lambda}, n))}{2E(Var(\hat{\lambda}, n))} \quad , \quad v = \frac{2[E(Var(\hat{\lambda}, n))]^2}{Var(Var(\hat{\lambda}, n))}$$

ซึ่งได้ทำการปรับปรุงช่วงความเชื่อมั่นสำหรับการประมาณค่าพารามิเตอร์ของการแจกแจงปัวซงด้วยวิธีการประมาณ 3 วิธี ดังนี้

1. วิธีประมาณแบบ Score (Score : SC)

ช่วงความเชื่อมั่น 100(1-α)% ของ λ โดยมีขอบเขตล่างและขอบเขตบนของช่วงความเชื่อมั่น

คือ $$\left(\hat{\lambda} + \frac{Z_{\frac{\alpha}{2}}^2}{2n} - \frac{Z_{\frac{\alpha}{2}}}{\sqrt{n}}\sqrt{\hat{\lambda} + \frac{Z_{\frac{\alpha}{2}}^2}{4n}},\ \hat{\lambda} + \frac{Z_{\frac{\alpha}{2}}^2}{2n} + \frac{Z_{\frac{\alpha}{2}}}{\sqrt{n}}\sqrt{\hat{\lambda} + \frac{Z_{\frac{\alpha}{2}}^2}{4n}}\right)$$

2. วิธีของ Wald แบบปรับค่าความต่อเนื่อง (Wald Interval with Continuity Correction : WCC)

ช่วงความเชื่อมั่น 100(1-α)% ของ λ โดยมีขอบเขตล่างและขอบเขตบนของช่วงความเชื่อมั่น

คือ $$\left(\hat{\lambda} - Z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{\lambda} + 0.5}{n}},\ \hat{\lambda} + Z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{\lambda} + 0.5}{n}}\right)$$

3. วิธีการแปลงข้อมูลที่ทำให้ความแปรปรวนคงที่ (Variance Stabilizing : VS)

ช่วงความเชื่อมั่น 100(1-α)% ของ λ โดยมีขอบเขตล่างและขอบเขตบนของช่วงความเชื่อมั่น

คือ

$$\left(\hat{\lambda}+\frac{\left(Z_{\frac{\alpha}{2}}\right)^2}{4n}-Z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{\lambda}+\frac{3}{8}}{n}},\ \hat{\lambda}+\frac{\left(Z_{\frac{\alpha}{2}}\right)^2}{4n}+Z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{\lambda}+\frac{3}{8}}{n}}\right)$$

จาก 3 วิธีการประมาณค่าพารามิเตอร์ดังกล่าว จะได้วิธีการประมาณช่วงความเชื่อมั่นด้วยการแจกแจงที ดังนี้

4. วิธีประมาณแบบ Score ด้วยการแจกแจงที (t - Scores : tSC)

วิธีการประมาณแบบ Score มีค่าความแปรปรวน คือ $Var\left(\frac{\hat{\lambda}+\frac{Z^2_{\frac{\alpha}{2}}}{4n}}{n}\right)$ นำค่าความแปรปรวนมาคำนวณค่าคาดหมายและค่าประมาณความแปรปรวน ดังนี้

$$E\left(\frac{\hat{\lambda}+\frac{t^2}{4n}}{n}\right)=\frac{1}{n}E\left(\hat{\lambda}+\frac{t^2}{4n}\right)=\frac{1}{n}\left(E\left(\hat{\lambda}\right)+E\left(\frac{t^2}{4n}\right)\right)$$

$$=\frac{1}{n}\left(\lambda+\frac{t^2}{4n}\right)=\frac{\lambda+\frac{t^2}{4n}}{n}$$

$$Var\left(\frac{\hat{\lambda}+\frac{t^2}{4n}}{n}\right)=\frac{1}{n^2}Var\left(\hat{\lambda}+\frac{t^2}{4n}\right)=\frac{1}{n^2}Var\left(\hat{\lambda}\right)$$

$$=\frac{1}{n^2}\left(\frac{\lambda}{n}\right)=\frac{\lambda}{n^3}$$

จากการประมาณค่าองศาเสรีของการแจกแจงของทีตามหลักการของ Satterthwaite’s method จะได้ค่าองศาเสรี

$$v=\frac{2\left[E\left(Var(\hat{\lambda})\right)\right]^2}{Var\left(Var(\hat{\lambda})\right)}=\frac{2\left(\frac{\lambda+\frac{t^2}{4n}}{n}\right)^2}{\frac{\lambda}{n^3}}=2\frac{n^3}{\lambda}\left(\frac{\lambda+\frac{t^2}{4n}}{n}\right)^2$$

$$=2\frac{n}{\lambda}\left(\lambda+\frac{t^2}{4n}\right)^2$$

ดังนั้นค่าของช่วงความเชื่อมั่น 100(1-α)% ของวิธีการประมาณแบบ Score ด้วยการแจกแจงที โดยมีขอบเขตล่างและขอบเขตบนของช่วงความเชื่อมั่น คือ

$$\left(\hat{\lambda}+\frac{t^2_{\frac{\alpha}{2}}}{2n}-\frac{t_{\frac{\alpha}{2}}}{\sqrt{n}}\sqrt{\hat{\lambda}+\frac{t^2_{\frac{\alpha}{2}}}{4n}},\ \hat{\lambda}+\frac{t^2_{\frac{\alpha}{2}}}{2n}+\frac{t_{\frac{\alpha}{2}}}{\sqrt{n}}\sqrt{\hat{\lambda}+\frac{t^2_{\frac{\alpha}{2}}}{4n}}\right)$$

ค่าองศาเสรี คือ $v=\frac{2n}{\hat{\lambda}}\left(\hat{\lambda}+\frac{t^2}{4n}\right)^2$

5. วิธีของ Wald แบบปรับค่าความต่อเนื่องด้วยการแจกแจงที (t - Wald Interval with Continuity Correction : tWCC)

วิธีของ Wald แบบปรับค่าความต่อเนื่อง มีค่าความแปรปรวนคือ $Var\left(\frac{\lambda+0.5}{n}\right)$ นำค่าความแปรปรวนมาคำนวณค่าคาดหมายและค่าประมาณความแปรปรวนดังนี้

$$E\left(\frac{\hat{\lambda}+0.5}{n}\right)=\frac{1}{n}E\left(\hat{\lambda}+0.5\right)=\frac{1}{n}\left(E\left(\hat{\lambda}\right)+E\left(0.5\right)\right)$$

$$=\frac{1}{n}\left(\lambda+0.5\right)=\frac{\lambda+0.5}{n}$$

$$Var\left(\frac{\hat{\lambda}+0.5}{n}\right)=\frac{1}{n^2}Var\left(\hat{\lambda}+0.5\right)=\frac{1}{n^2}\left(Var\left(\hat{\lambda}\right)+Var\left(0.5\right)\right)$$

$$=\frac{1}{n^2}\left(\frac{\lambda}{n}+0\right)=\frac{\lambda}{n^3}$$

จากการประมาณค่าองศาเสรีของการแจกแจงของทีตามหลักการของ Satterthwaite’s method ได้ค่าองศาเสรี คือ

$$v = \frac{2E\left[Var(\hat{\lambda})\right]^2}{Var\left[Var(\hat{\lambda})\right]} = \frac{2\left(\frac{\lambda+0.5}{n}\right)^2}{\frac{\lambda}{n^3}} = 2\frac{n^3}{\lambda}\left(\frac{\lambda+0.5}{n}\right)^2$$

$$= \frac{2n\left(\lambda+0.5\right)^2}{\lambda}$$

ดังนั้นช่วงความเชื่อมั่น 100(1-α)% ของวิธีของ Wald แบบปรับค่าความต่อเนื่องด้วยการแจกแจงที มีขอบเขตล่างและขอบเขตบนของช่วงความเชื่อมั่น

คือ $\left(\hat{\lambda} - t_{\frac{\alpha}{2},v}\sqrt{\frac{\hat{\lambda}+0.5}{n}},\ \hat{\lambda} + t_{\frac{\alpha}{2},v}\sqrt{\frac{\hat{\lambda}+0.5}{n}}\right)$ ค่าองศาเสรี

คือ $v = \dfrac{2n\left(\hat{\lambda}+0.5\right)^2}{\hat{\lambda}}$

6. วิธีการแปลงข้อมูลที่ทำให้ความแปรปรวนคงที่ด้วยการแจกแจงที (t - Variance Stabilizing : tVS)

จากวิธีการแปลงข้อมูลที่ทำให้ความแปรปรวนคงที่โดยวิธีของ Anscombe [Barker (2002)] ได้ค่าความแปรปรวนของวิธีนี้

คือ $Var\left(\frac{\hat{\lambda}+\frac{3}{8}}{n}\right)$ นำค่าความแปรปรวนมาคำนวณค่าคาดหมายและค่าประมาณความแปรปรวนดังนี้

$$E\left(\frac{\hat{\lambda}+\frac{3}{8}}{n}\right) = \frac{1}{n}E\left(\hat{\lambda}+\frac{3}{8}\right) = \frac{1}{n}\left(E(\hat{\lambda}) + E\left(\frac{3}{8}\right)\right)$$

$$= \frac{1}{n}\left(\lambda+\frac{3}{8}\right) = \frac{\lambda+\frac{3}{8}}{n}$$

$$Var\left(\frac{\hat{\lambda}+\frac{3}{8}}{n}\right) = \frac{1}{n^2}Var\left(\hat{\lambda}+\frac{3}{8}\right) = \frac{1}{n^2}Var(\hat{\lambda})$$

$$= \frac{1}{n^2}\left(\frac{\lambda}{n}\right) = \frac{\lambda}{n^3}$$

จากการประมาณค่าองศาเสรีของการแจกแจงของทีตามหลักการของ Satterthwaite's method จะได้ค่าองศาเสรี

$$v = \frac{2E\left[Var(\hat{\lambda})\right]^2}{Var\left[Var(\hat{\lambda})\right]} = \frac{2\left(\frac{\lambda+\frac{3}{8}}{n}\right)^2}{\frac{\lambda}{n^3}} = 2\frac{n^3}{\lambda}\left(\frac{\lambda+\frac{3}{8}}{n}\right)^2$$

$$= \frac{2n\left(\lambda+\frac{3}{8}\right)^2}{\lambda}$$

ดังนั้นค่าของช่วงความเชื่อมั่น 100(1-α)% ของวิธีการแปลงข้อมูลที่ทำให้ความแปรปรวนคงที่ ด้วยการแจกแจงทีมีขอบเขตล่างและขอบเขตบนของช่วงความเชื่อมั่น คือ

$$\left(\hat{\lambda} + \frac{t_{\frac{\alpha}{2}}^2}{4n} - t_{\frac{\alpha}{2}}\sqrt{\frac{\hat{\lambda}+\frac{3}{8}}{n}},\ \hat{\lambda} + \frac{t_{\frac{\alpha}{2}}^2}{4n} + t_{\frac{\alpha}{2}}\sqrt{\frac{\hat{\lambda}+\frac{3}{8}}{n}}\right)$$ ค่าองศาเสรี

คือ $v = \dfrac{2n\left(\lambda+\frac{3}{8}\right)^2}{\lambda}$

สำหรับแต่ละวิธีการประมาณช่วงความเชื่อมั่น (1-α)100% ของพารามิเตอร์ λ จะสามารถคำนวณหาค่าความน่าจะเป็นครอบคลุม (Coverage Probability) ดังนี้

$$CP(\lambda) = \sum_{i=1}^{n} I_{\{L(i)\leq\lambda\leq U(i)\}} \frac{e^{-\lambda}i^{-\lambda}}{i!}$$

เมื่อ I คือ ค่า Indicator จะนับ 1 เมื่อ $L\leq\lambda\leq U$

ที่ระดับความเชื่อมั่นที่กำหนดเท่ากับ 95% ความน่าจะเป็นของการครอบคลุมค่าพารามิเตอร์ λ ที่ระดับความเชื่อมั่นที่กำหนด

เมื่อ $0.95-1.960\sqrt{\frac{0.95(1-0.95)}{1000}} \leq CP \leq 0.95+1.960\sqrt{\frac{0.95(1-0.95)}{1000}}$

$$0.9360 \leq CP \leq 0.9640$$

ในทำนองเดียวกันค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่น (Expected Width of Confidence Interval) คำนวณดังนี้

$$EW(\lambda) = \sum_{i=1}^{n} [U(i)-L(i)] \frac{e^{-\lambda}\lambda^{i}}{i!}$$

เมื่อ $U(i)$คือ ขอบเขตบนของสถานการณ์ทำซ้ำรอบที่ i

$L(i)$ คือ ขอบเขตล่างของสถานการณ์ทำซ้ำรอบที่ i

ในการศึกษาวิธีการประมาณค่าแบบช่วงสำหรับพารามิเตอร์ λ ของการแจกแจงปัวซงครั้งนี้ มีขอบเขตการวิจัย ดังนี้

1. กำหนดขนาดตัวอย่างที่ใช้ คือ n = 10, 15, 30, 50 และ 100 ตามลำดับ ภายใต้การแจกแจงปัวซง

2. กำหนดค่าพารามิเตอร์ λ มีค่าเท่ากับ 1, 1.5, 3 และ 5 รวมทั้งหมด 4 ค่า

3. กำหนดระดับความเชื่อมั่น (1-α)100% เท่ากับ 95%

4. ศึกษาเปรียบเทียบการประมาณค่าแบบช่วงสำหรับพารามิเตอร์ λ ทั้ง 6 วิธี [Barker (2002)] คือ วิธีประมาณแบบ Score (SC) วิธีของ Wald แบบปรับค่าความต่อเนื่อง (WCC) วิธีการแปลงข้อมูลที่ทำให้ความแปรปรวนคงที่ (VS) วิธีประมาณแบบ Score ด้วยการแจกแจงที (tSC) วิธีของ Wald แบบปรับค่าความต่อเนื่องด้วยการแจกแจงที (tWCC) และวิธีการแปลงข้อมูลที่ทำให้ความแปรปรวนคงที่ด้วยการแจกแจงที (tVS)

5. กำหนดเกณฑ์ที่ใช้ในการเปรียบเทียบ เพื่อตรวจสอบค่าประมาณสัมประสิทธิ์ความเชื่อมั่น คือ ค่าความน่าจะเป็นครอบคลุม (CP) และค่าความกว้างเฉลี่ยของช่วง (EW) เพื่อหาวิธีการประมาณค่าพารามิเตอร์ที่ดีที่สุดในแต่ละสถานการณ์

6. ในการศึกษาครั้งนี้จะทำการจำลองข้อมูลซ้ำ 10,000 รอบในแต่ละสถานการณ์ของการทดลองด้วยโปรแกรม R เวอร์ชั่น 2.13.2

ผลการวิจัยและวิจารณ์ผล

ผลการตรวจสอบความน่าจะเป็นของการครอบคลุมค่าพารามิเตอร์ λ จากการจำลองข้อมูลภายใต้การแจกแจงปัวซง โดยกำหนด 1, 1.5, 3 และ 5 ขนาดตัวอย่าง n = 10, 15, 30, 50 และ 100 ที่ระดับความเชื่อมั่น 95% พบว่า

เมื่อพิจารณาค่าความน่าจะเป็นครอบคลุม วิธี SC พบว่ามีเพียงกรณีที่ขนาดตัวอย่าง n = 100, λ = 3, 5 เท่านั้นที่ให้ค่าครอบคลุมค่าพารามิเตอร์ที่กำหนด ส่วนวิธี tSC ที่ได้จากการปรับปรุงจากวิธี SC ให้ค่าความน่าจะเป็นครอบคลุมค่าพารามิเตอร์ λ ใกล้เคียงกับระดับความเชื่อมั่นที่กำหนดสูงสุดเกือบทุกกรณีโดยเฉพาะกรณีที่ขนาดตัวอย่างเล็กและกลาง

วิธี WCC ให้ค่าความน่าจะเป็นครอบคลุมค่าพารามิเตอร์เฉพาะกรณีที่ λ = 5 ของทุกขนาดตัวอย่าง เมื่อปรับค่าองศาเสรีของการแจกแจงทีแล้ว วิธี tWCC จะให้ค่าความน่าจะเป็นครอบคลุมระดับความเชื่อมั่นที่กำหนดใกล้เคียงหรือเท่ากับกับวิธีเดิมเป็นส่วนใหญ่

ส่วนวิธี VS ให้ค่าความน่าจะเป็นครอบคลุมในกรณีที่พารามิเตอร์ λ = 3, 5 เป็นส่วนใหญ่ และให้ค่าความน่าจะเป็นใกล้เคียงกับระดับความเชื่อมั่นที่กำหนด สำหรับตัวอย่างที่มีขนาดกลาง ขนาดตัวอย่าง $30 \leq n \leq 100$

จากผลการเปรียบเทียบค่าความกว้างเฉลี่ยพบว่า สำหรับตัวอย่างขนาดเล็ก n < 30 และตัวอย่างขนาดกลาง $30 \leq n \leq 100$ วิธี tSC ให้ค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่นที่น้อยที่สุดทุกกรณีศึกษา ซึ่งวิธีการประมาณค่าพารามิเตอร์ทุกวิธีให้ค่าความกว้างเฉลี่ยใกล้เคียงกัน ส่วนวิธี tWCC และวิธี tVS ให้ค่าความกว้างเฉลี่ยที่แคบขึ้นเล็กน้อยเมื่อเปรียบเทียบกับวิธีเดิมคือ WCC และ VS แต่ยังไม่ถือว่าเป็นวิธีที่ดีที่สุด เมื่อพิจารณาโดยรวมพบว่า ค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่นจะแคบลงเมื่อค่า n มีขนาดเพิ่มขึ้น ในทุกกรณีศึกษา ดังข้อมูลในตารางที่ 1

จากการจำลองข้อมูลภายใต้การแจกแจงปัวซงและทำการประมาณค่าช่วงความเชื่อมั่นด้วยวิธีการประมาณค่าทั้ง 3 วิธี คือ วิธีการประมาณค่าแบบ Score (SC) วิธีของ Wald แบบปรับค่าความต่อเนื่อง (WCC) วิธีการแปลงข้อมูลที่ทำให้ความแปรปรวนคงที่ (VS) โดยได้ทำการศึกษาแนวทางในการปรับปรุงช่วงความเชื่อมั่นที่ได้จากการประยุกต์ใช้วิธีการของ Satterthwaite’s method จากการศึกษาจากงานวิจัยของ Pan โดยประมาณเข้าสู่การแจกแจงที โดยทำการปรับค่าองศาเสรี คือ วิธีการประมาณค่าแบบ Score ด้วยการแจกแจงที (tSC) วิธีของ Wald แบบปรับค่าความต่อเนื่องด้วยการแจกแจงที (tWCC) วิธีการแปลงข้อมูลที่ทำให้ความแปรปรวนคงที่ด้วยการแจกแจงที (tVS) พบว่า จากการเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์แบบช่วงที่ระดับความเชื่อมั่น 95% วิธี tSC ให้ช่วงความเชื่อมั่นที่มีค่าความน่าจะเป็นครอบคลุมระดับความเชื่อมั่นที่กำหนดสูงที่สุดและให้ค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่นแคบที่สุดเมื่อเปรียบเทียบกับวิธีอื่นๆ ในสถานการณ์เดียวกัน ส่วนวิธี tWCC และ วิธี tVS เมื่อปรับค่าองศาเสรีโดยการแจกแจงทีแล้ว พบว่าค่าความน่าจะเป็นครอบคลุมค่าพารามิเตอร์ใกล้เคียงหรือเท่ากับวิธีเดิมเป็นส่วนใหญ่และให้ค่าความกว้างเฉลี่ยแคบกว่าวิธีเดิม เมื่อพิจารณาค่าความน่าจะเป็นครอบคลุมพบว่า ทุกวิธีเมื่อขนาดตัวอย่างและค่าพารามิเตอร์เพิ่มขึ้น จะมีค่าความน่าจะเป็นครอบคลุมมีแนวโน้มใกล้เคียงระดับความเชื่อมั่นที่กำหนดมากยิ่งขึ้น

ตารางที่ 1 ค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่นและค่าความน่าจะเป็นของการครอบคลุมพารามิเตอร์ λ จำแนกตามขนาดตัวอย่างและค่าพารามิเตอร์ λ ที่ระดับความเชื่อมั่น 95%

λ	n	Coverage Probability						Expected Width of Confidence Interval					
		S	WCC	VS	tSC	tWCC	tVS	S	WCC	VS	tSC	tWCC	tVS
1	10	0.9886	0.9681	0.9818	0.9285	0.9862	0.9818	1.6545	1.4708	1.4191	1.2339	1.5327	1.4658
	15	0.9827	0.9811	0.9731	0.9580	0.9811	0.9868	1.2742	1.2145	1.1506	1.0110	1.2365	1.1913
	30	0.9758	0.9841	0.9790	0.9565	0.9841	0.9790	0.8341	0.8620	0.8201	0.6994	0.8697	0.8289
	50	0.9710	0.9844	0.9813	0.9560	0.9844	0.9813	0.6182	0.6680	0.6372	0.5369	0.6716	0.6413
	100	0.9656	0.9844	0.9786	0.9451	0.9844	0.9813	0.4176	0.4726	0.4496	0.3733	0.4739	0.4524
1.5	10	0.9787	0.9731	0.9689	0.9544	0.9770	0.9689	1.9062	1.7031	1.6395	1.5124	1.7511	1.6830
	15	0.9726	0.9699	0.9607	0.9514	0.9699	0.9782	1.4810	1.3892	1.3297	1.2147	1.4105	1.3761
	30	0.9762	0.9762	0.9682	0.9485	0.9762	0.9682	0.9890	0.9878	0.9474	0.8433	0.9953	0.9556
	50	0.9687	0.9758	0.9749	0.9536	0.9758	0.9779	0.7369	0.7652	0.7393	0.6535	0.7687	0.7457
	100	0.9700	0.9750	0.9703	0.9528	0.9750	0.9703	0.5046	0.5407	0.5208	0.4596	0.5419	0.5221
3	10	0.9748	0.9572	0.9648	0.9572	0.9572	0.9648	2.4999	2.2219	2.1927	2.0990	2.2553	2.2285
	15	0.9769	0.9652	0.9549	0.9497	0.9652	0.9701	1.9792	1.8280	1.7730	1.6892	1.8462	1.8200
	30	0.9701	0.9701	0.9627	0.9490	0.9701	0.9627	1.3332	1.2991	1.2650	1.1851	1.3056	1.2717
	50	0.9672	0.9641	0.9604	0.9505	0.9641	0.9604	1.0066	1.0003	0.9783	0.9172	1.0033	0.9814
	100	0.9614	0.9649	0.9609	0.9479	0.9649	0.9638	0.6910	0.7077	0.6918	0.6449	0.7088	0.6951
5	10	0.9707	0.9541	0.9606	0.9541	0.9541	0.9606	3.0893	2.7759	2.7568	2.6789	2.8039	2.7861
	15	0.9697	0.9549	0.9599	0.9549	0.9632	0.9599	2.4585	2.2685	2.2514	2.1805	2.3016	2.2673
	30	0.9663	0.9593	0.9526	0.9514	0.9593	0.9574	1.6754	1.6108	1.5804	1.5295	1.6162	1.5933
	50	0.9619	0.9601	0.9603	0.9538	0.9601	0.9603	1.2692	1.2487	1.2342	1.1857	1.2512	1.2368
	100	0.9585	0.9604	0.9589	0.9535	0.9604	0.9589*	0.878	0.883	0.8715	0.8369	0.8839	0.8724

สรุปผลการวิจัย

เมื่อขนาดตัวอย่างเล็กและพารามิเตอร์มีค่าน้อย $n < 30$ การประมาณค่าของพารามิเตอร์ λ ควรเลือกใช้วิธี tSC เนื่องจากการประมาณค่าพารามิเตอร์ด้วยการแจกแจงทีให้ค่าความน่าจะเป็นครอบคลุมใกล้เคียงกับระดับความเชื่อมั่นที่กำหนดและให้ความกว้างเฉลี่ยช่วงความเชื่อมั่นน้อยที่สุด ซึ่งดีกว่าการประมาณวิธีเดิมด้วยการแจกแจงปรกติที่ใช้วิธี SC สำหรับขนาดตัวอย่างปานกลาง ($30 \leq n \leq 100$) และขนาดตัวอย่างที่ใหญ่ขึ้น ค่าพารามิเตอร์มีค่ามากขึ้น วิธี tSC จะให้ผลใกล้เคียงกันกับวิธี SC เนื่องจากเมื่อขนาดตัวอย่างใหญ่ การแจกแจงทีจะลู่เข้าสู่การแจกแจงปรกติ ทำให้วิธีการปรับปรุง ช่วงความเชื่อมั่นโดยประมาณด้วยการแจกแจงทีให้ค่าความน่าจะเป็นครอบคลุมใกล้เคียงหรือดีกว่าวิธีการประมาณค่าพารามิเตอร์วิธีเดิม ดังนั้นจึงสามารถนำวิธีการประมาณค่าด้วยวิธี tSC ไปปรับใช้ในการประมาณค่าพารามิเตอร์ในกรณีที่ประสบปัญหาข้อมูลที่มีการแจกแจงปัวซงที่ขนาดตัวอย่างน้อยได้ในทุกสถานการณ์ ซึ่งจะให้ความน่าจะเป็นครอบคลุมใกล้เคียงกับระดับความเชื่อมั่นที่กำหนดให้กว่าวิธีเดิมและได้ค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่นแคบลงด้วย

กิตติกรรมประกาศ

คณะผู้วิจัยขอขอบพระคุณผู้ประเมินบทความทุกท่านที่ให้คำแนะนำที่เป็นประโยชน์และข้อคิดดีๆ ในการเขียนงานวิจัยครั้งนี้

เอกสารอ้างอิง

Barker, L. (2002). A comparison of Nine Confidence Intervals for a Poisson Parameter When the Expected Number of Evens Is ≤ 5. *Journal of the American Statistical Association, 56*, 85-89.

Lehmann, E.L. (2005). *Testing Statistical Hypotheses*. New York, John Wiley & Sons. p.2-6.

R Development Core Team, (2009). *A Language and Environment for Statistical Computing*, R foundation for statistical computing, Vienna, Austria.

Pan, W. (2002). Approximate confidence intervals for one proportion and difference of two proportions. *Computational Statistics and Data Analysis, 40*, 143-157.