ORIGINAL ARTICLE

ช่วงความเชื่อมั่นสำหรับค่าเฉลี่ยของประชากรที่มีการแจกแจงแบบเบ้ขวา

วีรวรรณ ศักดาจิวะเจริญ[1] และ ธีระพร วีระถาวร[2]

Abstract

Sakdajivacharoen, W.[1] and Verathaworn, T.[2]
Confidence intervals for means of positively skewed distributions
Songklanakarin J. Sci. Technol., 2003, 25(4) : 485-496

The objective of this study is to compare interval estimation methods for population means of positively skewed distributions. The estimation methods are the interval estimation method with student-t statistics, the interval estimation method with Johnson's statistics, the interval estimation method with Hall's statistics and the interval estimation method with Chen's statistics. Log-normal distribution and Weibull distribution are considered. The measures of skewness under the consideration are 1.0, 3.0, 5.0, respectively. The sample sizes are 10, 30, 50 and the confidence levels are 0.95. The consideration has two steps. First, the confidence level of interval estimation methods are not lower than the determined confidence level value. The second is the comparision of mean of lower confidence limit, mean of upper confidence limit and mean of confidence interval length. The experimental data are generated by the Monte Carlo Simulation technique. The confidence level of interval estimation method with Bootstrap is higher than the non-boot-

[1]Department of Computer and Statistics, Faculty of Arts and Sciences, Dhurakijpundit University, Bangkok 10210 Thailand [2]Department of Statistics, Faculty of Commerce and Accountancy, Chulalongkorn University, Bangkok 10330 Thailand

[1]วท.ม.(สถิติ), ภาควิชาคอมพิวเตอร์และสถิติ คณะศิลปศาสตร์และวิทยาศาสตร์ มหาวิทยาลัยธุรกิจบัณฑิตย์ ประชาชื่น หลักสี่ กรุงเทพฯ 10210 [2]Ph.D.(Statistics), รองศาสตราจารย์, ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย ปทุมวัน กรุงเทพฯ 10330

Corresponding e-mail: weerawan_sakda@yahoo.com

รับต้นฉบับ 12 ธันวาคม 2545 รับลงพิมพ์ 30 เมษายน 2546

strap. The interval estimation method with Johnson's statistics is the optimum estimation method for the upper confidence interval and two-tailed confidence interval. The interval estimation method with Chen's statistics is the optimum estimation method for the lower confidence interval. Commonly, the confidence level of interval estimation methods for upper confidence interval are varied by the measure of skewness but the confidence level of interval estimation methods for lower confidence interval and two-tailed confidence interval are converted by the measure of skewness. The mean of lower confidence limit is varied by the sample size, on the other hand, the mean of upper confidence limit and mean of confidence interval length are converted by the sample size.

Key words : positively skewed distribution, confidence interval, bootstrap

บทคัดย่อ

วีรวรรณ ศักดาจิวะเจริญ และ ธีระพร วีระถาวร

ช่วงความเชื่อมั่นสำหรับค่าเฉลี่ยของประชากรที่มีการแจกแจงแบบเบ้ขวา

ว. สงขลานครินทร์ วทท. 2546 25(4) : 485-496

การวิจัยนี้มีจุดประสงค์เพื่อเปรียบเทียบวิธีการประมาณค่าแบบช่วงสำหรับค่าเฉลี่ยของประชากรที่มีการแจกแจงแบบเบ้ขวา โดยมีวิธีการประมาณค่าแบบช่วง 4 วิธี คือ วิธีการประมาณค่าแบบช่วงด้วยตัวสถิติที วิธีการประมาณค่าแบบช่วงด้วยตัวสถิติของจอห์นสัน วิธีการประมาณค่าแบบช่วงด้วยตัวสถิติของฮอลล์ และวิธีการประมาณค่าแบบช่วงด้วยตัวสถิติของเชน ซึ่งประชากรมีการแจกแจงลอกนอร์มอลและการแจกแจงไวบูลล์ ณ ระดับความเบ้ 1.0 3.0 5.0 โดยใช้ขนาดตัวอย่างเท่ากับ 10 30 50 และระดับความเชื่อมั่นเท่ากับ 0.95 เกณฑ์ที่ใช้ในการพิจารณาขั้นต้นได้พิจารณาว่าค่าระดับความเชื่อมั่นจากการทดลองที่ได้จากแต่ละวิธีการมีค่าไม่ต่ำกว่าที่กำหนด ขั้นตอนต่อไปได้พิจารณาเปรียบเทียบค่าเฉลี่ยของขีดจำกัดความเชื่อมั่นล่าง ค่าเฉลี่ยของขีดจำกัดความเชื่อมั่นบน และค่าเฉลี่ยความยาวของช่วงความเชื่อมั่น การวิจัยได้ใช้เทคนิคมอนติคาร์โล ผลสรุปของการวิจัยคือ ค่าระดับความเชื่อมั่นจากการทดลองกรณีที่ใช้บูตสแตรปในการหาช่วงความเชื่อมั่นมีค่าสูงกว่ากรณีที่ไม่ใช้บูตสแตรป วิธีการประมาณค่าแบบช่วงด้วยตัวสถิติของจอห์นสันเป็นวิธีการประมาณค่าแบบช่วงที่เหมาะสมสำหรับการทดสอบสมมติฐานทางเดียวด้านมากกว่าและการทดสอบสมมติฐานสองทาง ส่วนวิธีการประมาณค่าแบบช่วงด้วยตัวสถิติของเชนเป็นวิธีการประมาณค่าแบบช่วงที่เหมาะสมสำหรับการทดสอบสมมติฐานทางเดียวด้านน้อยกว่า โดยทั่วไป ค่าระดับความเชื่อมั่นจากการทดลองสำหรับการทดสอบสมมติฐานทางเดียวด้านน้อยกว่าของวิธีการประมาณค่าแบบช่วงทุกวิธีแปรผันตามค่าสัมประสิทธิ์ความเบ้ ส่วนค่าระดับความเชื่อมั่นจากการทดลองสำหรับการทดสอบสมมติฐานทางเดียวด้านมากกว่าและการทดสอบสมมติฐานสองทางของวิธีการประมาณค่าแบบช่วงทุกวิธีแปรผกผันกับค่าสัมประสิทธิ์ความเบ้ ค่าเฉลี่ยของขีดจำกัดความเชื่อมั่นล่างแปรผันตามขนาดตัวอย่าง ส่วนค่าเฉลี่ยของขีดจำกัดความเชื่อมั่นบนและค่าเฉลี่ยความยาวของช่วงความเชื่อมั่นแปรผกผันกับขนาดตัวอย่าง

การประมาณค่าแบบช่วงสำหรับค่าเฉลี่ยของประชากร โดยทั่วไปจะใช้ตัวสถิติ Z (Z-statistic) ในการสร้างช่วงความเชื่อมั่นสำหรับค่าเฉลี่ยของประชากรเมื่อทราบส่วนเบี่ยงเบนมาตรฐานของประชากร และตัวสถิติที (T-statistic) ในการสร้างช่วงความเชื่อมั่นสำหรับค่าเฉลี่ยของประชากรเมื่อไม่ทราบส่วนเบี่ยงเบนมาตรฐานของประชากร การใช้ตัวสถิติดังกล่าวข้างต้นนี้จะอยู่ภายใต้ข้อตกลงเบื้องต้น คือ ประชากรมีการแจกแจงแบบปกติ บางครั้งในทางปฏิบัติประชากรที่สนใจศึกษาอาจเป็นการแจกแจงอื่นที่ไม่ใช่การแจกแจงแบบปกติ อย่างไรก็ตาม กรณีที่ขนาดตัวอย่างมีขนาดใหญ่ก็จะสามารถประมาณได้ว่าข้อมูลที่ได้มีการแจกแจงใกล้เคียงกับการแจกแจงแบบ

ปกติโดยอาศัยทฤษฎีค่าจำกัดสู่ส่วนกลาง (The Central Limit Theorem) แต่ในกรณีที่ขนาดตัวอย่างมีขนาดเล็ก การใช้ตัวสถิติดังกล่าวในการสร้างช่วงความเชื่อมั่นสำหรับค่าเฉลี่ยอาจจะทำให้ผลสรุปที่ได้มีความผิดพลาด

ในปี ค.ศ.1928 เนย์แมน (Neyman) และเพียร์-สัน (Pearson) ได้แสดงให้เห็นถึงผลกระทบจากการใช้ตัวสถิติที เมื่อประชากรไม่ได้มีการแจกแจงเป็นแบบปกติและขนาดตัวอย่างมีขนาดเล็ก เขาพบว่าความเบ้ของประชากรจะมีผลต่อการแจกแจงของตัวสถิติทีมากกว่าความโด่ง และประชากรที่มีลักษณะเบ้ขวามีผลกระทบต่อการแจกแจงของตัวสถิติทีในทางเบ้ซ้าย ในทางกลับกัน ประชากรที่มีลักษณะเบ้ซ้ายจะมีผลกระทบต่อการแจกแจงของตัวสถิติทีในทางเบ้ขวา ในปี ค.ศ.1978 จอห์นสัน (Johnson) ได้เสนอตัวสถิติ modified t ซึ่งได้ทำการแปลงจากตัวสถิติทีเพื่อให้มีความเหมาะสมสำหรับการประมาณค่าเฉลี่ยของการแจกแจงที่ไม่สมมาตรโดยใช้ Cornish-Fisher Expansion ต่อมาในปี ค.ศ.1992 ฮอลล์ (Hall) ได้เสนอวิธีการแปลงตัวสถิติทีด้วย Edgeworth Expansion ในปี ค.ศ. 1995 เชน (Chen) ได้เสนอตัวสถิติทดสอบใหม่ซึ่งทำการแปลงจากตัวสถิติทีโดยใช้ Edgeworth Expansion และ Taylor Expansion หลังจากนั้นในปี ค.ศ.2000 ซู (Zhou) และเกาว์ (Gao) ได้แสดงให้เห็นว่าการนำวิธีการบูตสแตรป (Bootstrap method) มาช่วยในการหาค่าของตัวสถิติของจอห์นสันและฮอลล์จะทำให้ผลที่ได้มีความแม่นยำมากยิ่งขึ้น

ดังนั้นในงานวิจัยนี้ผู้วิจัยจึงสนใจที่จะทำการศึกษาเพื่อเปรียบเทียบวิธีการประมาณค่าแบบช่วงสำหรับค่าเฉลี่ยของประชากรซึ่งมีการแจกแจงแบบเบ้ขวาที่ระดับความเบ้ต่างกันในการแจกแจงหนึ่งๆ จำแนกตามขนาดตัวอย่างและระดับความเชื่อมั่น โดยพิจารณาวิธีการประมาณค่าแบบช่วง ดังต่อไปนี้

1. วิธีการประมาณค่าแบบช่วงด้วยตัวสถิติที (T)
2. วิธีการประมาณค่าแบบช่วงด้วยตัวสถิติของจอห์นสัน (J)
3. วิธีการประมาณค่าแบบช่วงด้วยตัวสถิติของฮอลล์ (H)
4. วิธีการประมาณค่าแบบช่วงด้วยตัวสถิติของเชน (C)

วิธีการประมาณค่าแบบช่วงที่ใช้ในการวิจัย

การประมาณค่าแบบช่วงสำหรับค่าเฉลี่ยของประชากรในกรณีที่ไม่ทราบค่าส่วนเบี่ยงเบนมาตรฐานของประชากรนั้นจะใช้ตัวสถิติทีในการสร้างช่วงความเชื่อมั่น ตัวสถิติทีนี้จะมีการแจกแจงทีด้วย ร.ส. (ระดับขั้นความเสรี (degree of freedom)) เท่ากับ n-1 ซึ่งตัวสถิติทีจะอยู่ในรูปของ

$$T = \frac{\bar{X} - \mu}{S / \sqrt{n}}$$

เมื่อ $\bar{X}$ แทนค่าเฉลี่ยของตัวอย่างเท่ากับ $\frac{\sum_{i=1}^{n} X_i}{n}$

S แทนส่วนเบี่ยงเบนมาตรฐานของตัวอย่าง เท่ากับ $\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2}$

และ n แทนขนาดตัวอย่าง

ดังนั้นช่วงความเชื่อมั่นด้านบน (1-α)100% ของค่าเฉลี่ยสำหรับการทดสอบสมมติฐานทางด้านมากกว่า คือ $(0, \bar{X} + t_{1-\alpha,(n-1)}\frac{S}{\sqrt{n}})$ ส่วนช่วงความเชื่อมั่นด้านล่าง (1-α) 100% ของค่าเฉลี่ยสำหรับการทดสอบสมมติฐานทางด้านน้อยกว่า คือ $(\bar{X} - t_{1-\alpha,(n-1)}\frac{S}{\sqrt{n}}, \infty)$ และช่วงความเชื่อมั่น (1-α)100% ของค่าเฉลี่ยสำหรับการทดสอบสมมติฐานสองทาง คือ $(\bar{X} - t_{1-\alpha/2,(n-1)}\frac{S}{\sqrt{n}}, \bar{X} + t_{1-\alpha/2,(n-1)}\frac{S}{\sqrt{n}})$

ในปี ค.ศ.1978 จอห์นสัน (Johnson) ได้ทำการแปลงตัวสถิติทีเพื่อให้เหมาะสมสำหรับการประมาณค่าเฉลี่ยของการแจกแจงที่ไม่สมมาตรโดยอาศัยหลักการของ Cornish-Fisher Expansion ซึ่งจะแทนที่ $\bar{X} - \mu$ ในตัวแปรทีด้วยสองเทอมแรกของ Inverse Cornish-Fisher Expansion[1] ดังนั้นตัวสถิติของจอห์นสันจะอยู่ในรูปของ

$$g_1(t) = t + \frac{\hat{\gamma}}{\sqrt{n}}\left(\frac{t^2}{3} + \frac{1}{6}\right)$$

[1]ฟังก์ชัน inverse ของ cornish fisher expansion

เมื่อ $\hat{\gamma}$ แทนสัมประสิทธิ์ความเบ้ซึ่งมีค่าคือเท่ากับ

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^3/S^3$$

จะได้ว่า

$$g_1^{-1}(z) = \left[-1\pm\sqrt{1-4\left(\frac{\hat{\gamma}}{3\sqrt{n}}\right)\left(\frac{\hat{\gamma}}{6\sqrt{n}} - z\right)}\right] / \left[2\left(\frac{\hat{\gamma}}{3\sqrt{n}}\right)\right]$$

ซึ่งจะหาค่าได้ก็ต่อเมื่อ

$$1-4\left(\frac{\hat{\gamma}}{3\sqrt{n}}\right)\left(\frac{\hat{\gamma}}{6\sqrt{n}} - z\right) \geq 0$$

จากการแก้สมการจะได้ค่า $g_1^{-1}(z)$ สองค่า แต่ในที่นี้จะใช้ค่า

$$g_1^{-1}(z) = \left[-1+\sqrt{1-4\left(\frac{\hat{\gamma}}{3\sqrt{n}}\right)\left(\frac{\hat{\gamma}}{6\sqrt{n}} - z\right)}\right] / \left[2\left(\frac{\hat{\gamma}}{3\sqrt{n}}\right)\right]$$

เพราะเป็นคำตอบของสมการที่อยู่ในช่วงของฟังก์ชันเพิ่ม (ตัวสถิติของจอห์นสันเป็นการแปลงตัวสถิติทีเพื่อหาค่าเปอร์เซนต์ไทล์ที่ (1-α)100 ของการแจกแจงที่สมมาตร ซึ่งค่าเปอร์เซนต์ไทล์นี้มีลักษณะเป็นฟังก์ชันเพิ่ม)

ดังนั้นช่วงความเชื่อมั่นด้านบน (1-α)100% ของค่าเฉลี่ยสำหรับการทดสอบสมมติฐานทางด้านมากกว่า คือ $(0, \bar{X} - g_1^{-1}(z_\alpha)\frac{S}{\sqrt{n}})$ ส่วนช่วงความเชื่อมั่นด้านล่าง (1-α) 100% ของค่าเฉลี่ยสำหรับการทดสอบสมมติฐานทางด้านน้อยกว่า คือ $(\bar{X} - g_1^{-1}(z_{1-\alpha})\frac{S}{\sqrt{n}}, \infty)$ และช่วงความเชื่อมั่น (1-α)100% ของค่าเฉลี่ยสำหรับการทดสอบสมมติฐานสองทาง คือ $(\bar{X} - g_1^{-1}(z_{1-\alpha/2})\frac{S}{\sqrt{n}}, \bar{X} - g_1^{-1}(z_{\alpha/2})\frac{S}{\sqrt{n}})$

ในปี ค.ศ.1992 ฮอลล์ (Hall) ได้ทำการแปลงตัวสถิติทีเพื่อให้มีความเหมาะสมสำหรับการประมาณค่าเฉลี่ยของการแจกแจงที่ไม่สมมาตรเช่นเดียวกัน แต่อาศัยหลักการของ Edgeworth Expansion จะได้ว่า ตัวสถิติของฮอลล์จะอยู่ในรูปของ

$$g_2(t) = t + \frac{\hat{\gamma}}{\sqrt{n}}\left(\frac{t^2}{3} + \frac{1}{6}\right) + \frac{\hat{\gamma}^2 t^3}{27n}$$

และ $$g_2^{-1}(z) = \frac{3\sqrt{n}}{\hat{\gamma}}\left[\left\{1+\hat{\gamma}\left(\frac{z}{\sqrt{n}} - \frac{\hat{\gamma}}{6n}\right)\right\}^{1/3} -1\right]$$

ดังนั้นช่วงความเชื่อมั่นด้านบน (1-α)100% ของค่าเฉลี่ยสำหรับการทดสอบสมมติฐานทางด้านมากกว่า คือ $(0, \bar{X} - g_2^{-1}(z_\alpha)\frac{S}{\sqrt{n}})$ ส่วนช่วงความเชื่อมั่นด้านล่าง (1-α) 100% ของค่าเฉลี่ยสำหรับการทดสอบสมมติฐานทางด้านน้อยกว่า คือ $(\bar{X} - g_2^{-1}(z_{1-\alpha})\frac{S}{\sqrt{n}}, \infty)$ และช่วงความเชื่อมั่น (1-α)100% ของค่าเฉลี่ยสำหรับการทดสอบสมมติฐานสองทาง คือ $(\bar{X} - g_2^{-1}(z_{1-\alpha/2})\frac{S}{\sqrt{n}}, \bar{X} - g_2^{-1}(z_{\alpha/2})\frac{S}{\sqrt{n}})$

ในปี ค.ศ.1995 เชน (Chen) ได้ทำการแปลงตัวสถิติทีเพื่อให้มีความเหมาะสมสำหรับการประมาณค่าเฉลี่ยของการแจกแจงแบบเบ้ขวาด้วยกลุ่มตัวอย่างขนาดเล็กโดยอาศัยหลักการของ Edgeworth Expansion และ Taylor Expansion จะได้ว่า ตัวสถิติของเชนจะอยู่ในรูปของ

$$g_3(t) = t + \frac{\hat{\gamma}}{\sqrt{n}}\left(\frac{t^2}{3} + \frac{1}{6}\right) + \frac{\hat{\gamma}^2}{9n}\left(t + 2t^3\right)$$

สำหรับการหา $g_3^{-1}(z)$ จะใช้การวิเคราะห์เชิงตัวเลข (numerical analysis) โดยจะใช้วิธีของนิวตัน-ราฟสัน (Newton-Raphson method)

ดังนั้นช่วงความเชื่อมั่นด้านบน (1-α)100% ของค่าเฉลี่ยสำหรับการทดสอบสมมติฐานทางด้านมากกว่า คือ $(0, \bar{X} - g_3^{-1}(z_\alpha)\frac{S}{\sqrt{n}})$ ส่วนช่วงความเชื่อมั่นด้านล่าง (1-α) 100% ของค่าเฉลี่ยสำหรับการทดสอบสมมติฐานทางด้านน้อยกว่า คือ $(\bar{X} - g_3^{-1}(z_{1-\alpha})\frac{S}{\sqrt{n}}, \infty)$ และช่วงความเชื่อมั่น (1-α)100% ของค่าเฉลี่ยสำหรับการทดสอบสมมติฐานสองทาง คือ $(\bar{X} - g_3^{-1}(z_{1-\alpha/2})\frac{S}{\sqrt{n}}, \bar{X} - g_3^{-1}(z_{\alpha/2})\frac{S}{\sqrt{n}})$

สำหรับการวิจัยนี้ ผู้วิจัยได้นำวิธีการบูตสแตรปมาช่วยในการหาค่าเปอร์เซนต์ไทล์ที่ $\alpha\times100$ ของการแจกแจงของ $g(t)$ ซึ่งเป็นฟังก์ชันการแปลงตัวสถิติทีของวิธีการประมาณค่าแบบช่วงต่างๆ ข้างต้น สำหรับการสร้างช่วงความเชื่อมั่น โดยกำหนดให้ $X^* = (X_1^*, ..., X_n^*)'$ แทนตัวอย่างสุ่มขนาด n ซึ่งสุ่มแบบใส่คืนจากตัวอย่างชุดแรก $X = (X_1, ..., X_n)'$ ให้ $g^*(t^*)$ เป็นการแจกแจงแบบมีเงื่อนไขบน X เรียกว่าการแจกแจงบูตสแตรปของ $g(t)$ ซึ่งคำนวณจากตัวอย่างซ้ำ X^* โดยที่การแจกแจงบูตสแตรปของ $g(t)$ สำหรับตัวอย่างชุดที่ b คือ $g^*(t^*)_b$ และให้ $\hat{t}^{(\alpha)}$

แทนเปอร์เซนต์ไทล์ที่ 100α ของการแจกแจงบูตสแตรปของ $g(t)$ ซึ่งพิจารณาจาก $\hat{t}^{(\alpha)}$ จะเป็นค่าที่ $B\alpha$ ของ $g^*(t^*)_b$

ดังนั้นช่วงความเชื่อมั่นด้านบน $(1-\alpha)100\%$ ของค่าเฉลี่ยสำหรับการทดสอบสมมติฐานทางด้านมากกว่า คือ $(0, \bar{X} - g^{-1}(\hat{t}^{(\alpha)})\frac{S}{\sqrt{n}})$ ส่วนช่วงความเชื่อมั่นด้านล่าง $(1-\alpha)100\%$ ของค่าเฉลี่ยสำหรับการทดสอบสมมติฐานทางด้านน้อยกว่า คือ $(\bar{X} - g^{-1}(\hat{t}^{(1-\alpha)})\frac{S}{\sqrt{n}}, \infty)$ และช่วงความเชื่อมั่น $(1-\alpha)100\%$ ของค่าเฉลี่ยสำหรับการทดสอบสมมติฐานสองทาง คือ

$$(\bar{X} - g^{-1}(\hat{t}^{(1-\alpha/2)})\frac{S}{\sqrt{n}}, \bar{X} - g^{-1}(\hat{t}^{(\alpha/2)})\frac{S}{\sqrt{n}})$$

การจำลองข้อมูลและผลการวิจัย

ในการศึกษาครั้งนี้กำหนดสถานการณ์ต่างๆ ดังนี้ การแจกแจงของประชากร ได้แก่ การแจกแจงลอกนอร์มอลและการแจกแจงไวบูลล์ สัมประสิทธิ์ความเบ้ของการแจกแจงต่างๆ เท่ากับ 1.0, 3.0 และ 5.0 ขนาดตัวอย่าง เท่ากับ 10, 30 และ 50 ค่าระดับความเชื่อมั่น เท่ากับ 0.95 การจำลองข้อมูลจะใช้เทคนิควิธีการจำลองแบบมอนติคาร์โลโดยที่ค่าระดับความเชื่อมั่นจากการทดลองจะคำนวณจาก

$$\frac{\text{จำนวนครั้งที่ช่วงความเชื่อมั่นครอบคลุมค่าพารามิเตอร์ที่กำหนด}}{\text{จำนวนครั้งที่ทำการศึกษาทั้งหมด 3000 ครั้ง}}$$

ในการพิจารณาจะพิจารณาแยกเป็น 2 ขั้นตอน ในขั้นตอนแรกจะพิจารณาเปรียบเทียบค่าระดับความเชื่อมั่นจากการทดลองที่ได้จากแต่ละวิธีการว่าเท่ากับค่าระดับความเชื่อมั่นที่กำหนดหรือไม่ โดยอาศัยการทดสอบสมมติฐาน

$H_0 : p \leq 0.95$

เทียบกับ $H_1 : p > 0.95$ จะได้ว่า

เราจะปฏิเสธสมมติฐาน H_0 เมื่อ

$$\hat{p} > 0.95 + 1.645\sqrt{(0.95 \times 0.05) / 3000}$$

$$\hat{p} > 0.9565$$

เพราะฉะนั้น ถ้าวิธีการใดให้ค่าระดับความเชื่อมั่นจากการทดลองไม่ต่ำกว่า 0.9565 แสดงว่าวิธีการนั้นให้ค่าระดับความเชื่อมั่นตามค่าที่กำหนด

ในขั้นตอนที่สอง ถ้าค่าระดับความเชื่อมั่นจากการทดลองของวิธีการใดมีค่าเท่ากับระดับความเชื่อมั่นที่กำหนด จะพิจารณาเปรียบเทียบค่าเฉลี่ยของขีดจำกัดความเชื่อมั่น

Table 1. The confidence level of interval estimation methods at the determined confidence level 0.95 from log-normal population. (lower confidence interval)

Skewness Coefficient	**1.0**			**3.0**			**5.0**		
Sample size	**10**	**30**	**50**	**10**	**30**	**50**	**10**	**30**	**50**
STUDENT t	0.9687 (0.869)	0.9640 (0.950)	0.9617 (0.973)	0.9840 (0.808)	0.9733 (1.019)	0.9697 (1.078)	0.9867 (0.811)	0.9737 (1.115)	0.9727 (1.024)
JOHNSON	0.9957 (0.910)	0.9977 (0.964)	0.9977 (0.982)	0.9973 (0.940)	0.9973 (1.069)	0.9973 (1.109)	0.9990 (1.000)	0.9993 (1.195)	0.9997 (1.255)
HALL	0.9957 (0.914)	0.9977 (0.964)	0.9977 (0.982)	0.9973 (0.962)	0.9973 (1.071)	0.9977 (1.110)	0.9990 (1.047)	0.9993 (1.201)	0.9993 (1.256)
CHEN	0.9953 (0.916)*	0.9977 (0.965)*	0.9977 (0.983)*	0.9853 (0.969)*	0.9967 (1.075)*	0.9973 (1.112)*	0.9710 (1.080)*	0.9987 (1.209)*	0.9993 (1.261)*

Data in the parentheses are mean of lower confidence limit in the case that the confidence coefficient of interval estimation method is not lower than the determined confidence coefficient value.
*** is the maximum MLCL in any case.**

Table 2. The confidence level of interval estimation methods at the determined confidence level 0.95 from log-normal population. (upper confidence interval)

Skewness Coefficient	**1.0**			**3.0**			**5.0**		
Sample size	**10**	**30**	**50**	**10**	**30**	**50**	**10**	**30**	**50**
STUDENT t	0.9113	0.9223	0.9297	0.8423	0.8853	0.9027	0.8117	0.8603	0.8787
JOHNSON	0.9773 (1.205)*	0.9923 (1.156)*	0.9943 (1.131)*	0.9380	0.9870 (1.606)*	0.9910 (1.551)*	0.9247	0.9790 (2.012)*	0.9877 (1.933)*
HALL	0.9793 (1.209)	0.9927 (1.156)	0.9943 (1.131)	0.9500	0.9893 (1.610)	0.9930 (1.551)	0.9390	0.9903 (2.026)	0.9917 (1.938)
CHEN	0.9937 (1.238)	0.9943 (1.157)	0.9963 (1.131)	0.9950 (1.872)*	0.9957 (1.630)	0.9983 (1.555)	0.9943 (2.485)*	0.9967 (2.082)	0.9993 (1.959)

Data in the parentheses are mean of upper confidence limit in the case that the confidence coefficient of interval estimation method is not lower than the determined confidence coefficient value.
*** is the minimum MUCL in any case.**

Table 3. The confidence level of interval estimation methods at the determined confidence level 0.95 from log-normal population. (two-tailed confidence interval)

Skewness Coefficient	**1.0**			**3.0**			**5.0**		
Sample size	**10**	**30**	**50**	**10**	**30**	**50**	**10**	**30**	**50**
STUDENT t	0.9383	0.9417	0.9417	0.8803	0.9123	0.9160	0.8503	0.8877	0.9070
JOHNSON	0.9733 (0.352)*	0.9907 (0.228)*	0.9947 (0.175)*	0.9370	0.9847 (0.655)*	0.9893 (0.532)*	0.9263	0.9793 (2.012)*	0.9870 (1.933)*
HALL	0.9783 (0.355)	0.9940 (0.229)	0.9950 (0.176)	0.9567 (0.942)*	0.9900 (0.655)	0.9920 (0.536)	0.9500	0.9890 (2.026)	0.9917 (1.939)
CHEN	0.9920 (0.413)	0.9957 (0.229)	0.9973 (0.176)	0.9880 (1.161)	0.9933 (0.675)	0.9957 (0.536)	0.9743 (1.841)*	0.9927 (2.082)	0.9940 (1.959)

Data in the parentheses are mean of upper confidence limit in the case that of the confidence coefficient of interval estimation method is not lower than the determined confidence coefficient value.
*** is the minimum MCIL in any case.**

ล่างจากการทดสอบสมมติฐานทางด้านน้อยกว่า (MLCL) ค่าเฉลี่ยของขีดจำกัดความเชื่อมั่นบนจากการทดสอบสมมติฐานทางด้านมากกว่า (MUCL) และค่าเฉลี่ยความยาวของช่วงความเชื่อมั่นจากการทดสอบสมมติฐานสองทาง (MCIL) ในการศึกษาครั้งนี้ทำการทดลองซ้ำ 3,000 ครั้ง ซึ่งแต่ละครั้งมีการกระทำจำนวนรอบของวิธีบูตสแตรปเท่ากับ 2,000 ครั้ง

จาก Table ที่ 1-6 ข้างต้นในกรณีของการทดสอบสมมติฐานทางเดียวด้านน้อยกว่า วิธีการประมาณค่าแบบช่วงทุกวิธีให้ค่าระดับความเชื่อมั่นจากการทดลองไม่ต่ำกว่าที่กำหนดสำหรับทุกค่าสัมประสิทธิ์ความเบ้และขนาดตัวอย่าง วิธี C ให้ค่า MLCL มากที่สุดสำหรับทุกขนาดตัวอย่าง

Table 4. The confidence level of interval estimation methods at the determined confidence level 0.95 from weibull population. (lower confidence interval)

Skewness Coefficient	1.0			3.0			5.0		
Sample size	**10**	**30**	**50**	**10**	**30**	**50**	**10**	**30**	**50**
STUDENT t	0.9747 (0.591)	0.9633 (0.725)	0.9600 (0.765)	0.9850 (0.503)	0.9753 (0.790)	0.9720 (0.870)	0.9973 (0.529)	0.9757 (0.966)	0.9730 (1.098)
JOHNSON	0.9963 (0.655)	0.9970 (0.751)	0.9970 (0.780)	0.9987 (0.663)	0.9987 (0.849)	0.9997 (0.906)	0.9980 (0.868)	0.9993 (1.051)	1.0000 (1.152)
HALL	0.9963 (0.662)	0.9970 (0.751)	0.9973 (0.780)	0.9980 (0.705)	0.9987 (0.854)	0.9997 (0.907)	0.9987 (1.006)	0.9990 (1.071)	1.0000 (1.160)
CHEN	0.9947 (0.666)*	0.9967 (0.752)*	0.9980 (0.781)*	0.9793 (0.728)*	0.9973 (0.861)*	0.9977 (0.911)*	0.9313 (1.092)*	0.9940 (1.093)*	0.9973 (1.172)*

Data in the parentheses are mean of lower confidence limit in the case that the confidence coefficient of interval estimation method is not lower than the determined confidence coefficient value.
*** is the maximum MLCL in any case.**

Table 5. The confidence level of interval estimation methods at the determined confidence level 0.95 from weibull population. (upper confidence interval)

Skewness Coefficient	1.0			3.0			5.0		
Sample size	**10**	**30**	**50**	**10**	**30**	**50**	**10**	**30**	**50**
STUDENT t	0.9073	0.9280	0.9317	0.8493	0.9079	0.9180	0.6553	0.8977	0.9040
JOHNSON	0.9773 (1.173)*	0.9957 (1.083)*	0.9987 (1.041)*	0.9470	0.9903 (1.635)*	0.9947 (1.556)*	0.8430	0.9793 (2.435)*	0.9900 (2.298)*
HALL	0.9813 (1.181)	0.9963 (1.084)	0.9987 (1.041)	0.9660 (1.827)*	0.9957 (1.651)	0.9973 (1.564)	0.8937	0.9950 (2.498)	0.9973 (2.328)
CHEN	0.9963 (1.249)	0.9977 (1.088)	0.990 (1.042)	0.9980 (2.114)	0.9997 (1.713)	1.0000 (1.584)	0.9987 (2.875)*	0.9997 (2.685)	1.0000 (2.417)

Data in the parentheses are mean of upper confidence limit in the case that the confidence coefficient of interval estimation method is not lower than the determined confidence coefficient value.
*** is the minimum MUCL in any case.**

รองลงมาคือ วิธี H วิธี J และวิธี T ตามลำดับ กรณีของสมมติฐานทางเดียวด้านมากกว่า วิธี T ให้ค่าระดับความเชื่อมั่นจากการทดลองต่ำกว่าที่กำหนดสำหรับทุกค่าสัมประสิทธิ์ความเบ้และขนาดตัวอย่าง ส่วนวิธีการประมาณค่าแบบช่วงอื่นๆ ส่วนใหญ่ให้ค่าระดับความเชื่อมั่นจากการทดลองไม่ต่ำกว่าที่กำหนด วิธี J ให้ค่า MUCL น้อยที่สุดสำหรับทุกค่าสัมประสิทธิ์ความเบ้และขนาดตัวอย่าง รองลงมาคือ วิธี H และวิธี C ตามลำดับ กรณีสมมติฐานสองทาง วิธี T ให้ค่าระดับความเชื่อมั่นจากการทดลองต่ำกว่าที่กำหนดสำหรับทุกค่าสัมประสิทธิ์ความเบ้และขนาดตัวอย่าง

Table 6. The confidence level of interval estimation methods at the determined confidence level 0.95 from weibull population. (two-tailed confidence interval)

Skewness Coefficient	1.0			3.0			5.0		
Sample size	**10**	**30**	**50**	**10**	**30**	**50**	**10**	**30**	**50**
STUDENT t	0.9363	0.9407	0.9467	0.8853	0.9260	0.9327	0.7247	0.9187	0.9227
JOHNSON	0.9747 (0.615)*	0.9933 (0.397)*	0.9963 (0.309)*	0.9473	0.9897 (0.964)*	0.9947 (0.794)*	0.8463	0.9793 (1.716)*	0.9900 (1.414)*
HALL	0.9793 (0.625)	0.9967 (0.397)	0.9983 (0.309)	0.9723 (1.514)*	0.9940 (0.977)	0.9953 (0.796)	0.9167	0.9937 (1.842)	0.9950 (1.445)
CHEN	0.9940 (0.757)	0.9977 (0.401)	0.9980 (0.309)	0.9810 (1.832)	0.9953 (1.054)	0.9977 (0.818)	0.9370	0.9943 (2.025)	0.9973 (1.546)

Data in the parentheses are mean of upper confidence limit in the case that the confidence coefficient of interval estimation method is not lower than the determined confidence coefficient value.
*** is the minimum MCIL in any case.**

ส่วนวิธีอื่นๆ ส่วนใหญ่ให้ค่าระดับความเชื่อมั่นจากการทดลองไม่ต่ำกว่าที่กำหนด วิธี J ให้ค่า MCIL น้อยที่สุดสำหรับทุกค่าสัมประสิทธิ์ความเบ้และขนาดตัวอย่าง รองลงมาคือ วิธี H และวิธี C ตามลำดับ โดยทั่วไป ค่าระดับความเชื่อมั่นจากการทดลองสำหรับการทดสอบสมมติฐานทางเดียวด้านน้อยกว่าของวิธีการประมาณค่าแบบช่วงทุกวิธีแปรผันตามกับค่าสัมประสิทธิ์ความเบ้ ส่วนค่าระดับความเชื่อมั่นจากการทดลองสำหรับการทดสอบสมมติฐานทางเดียวด้านมากกว่าและการทดสอบสมมติฐานสองทางของวิธีการประมาณค่าแบบช่วงทุกวิธีแปรผกผันกับค่าสัมประสิทธิ์ความเบ้ ค่าเฉลี่ยของขีดจำกัดความเชื่อมั่นล่างแปรผันตามขนาดตัวอย่าง ส่วนค่าเฉลี่ยของขีดจำกัดความ

Table 7. The optimum estimation method for the confidence interval at th determined confidence level 0.95 of any populations and measure of skewness. The sample sizes are 10, 30 and 50 respectively.

Measure of skewness	Sample sizes	$H_0 : \mu = \mu_o$ $H_0 : \mu < \mu_o$		$H_0 : \mu = \mu_o$ $H_0 : \mu > \mu_o$		$H_0 : \mu = \mu_o$ $H_0 : \mu \neq \mu_o$	
		Log-normal	**Weibull**	**Log-normal**	**Weibull**	**Log-normal**	**Weibull**
1.0	10	C	C	J	J	J	J
	30	C	C	J	J	J	J
	50	C	C	J	J	J	J
3.0	10	C	C	C	H	H	H
	30	C	C	J	J	J	J
	50	C	C	J	J	J	J
5.0	10	C	C	C	C	C	-
	30	C	C	J	J	J	J
	50	C	C	J	J	J	J

- There is no optimum estimation method.

เชื่อมั่นบนและค่าเฉลี่ยความยาวของช่วงความเชื่อมั่นแปรผกผันกับขนาดตัวอย่าง ในกรณีที่ไม่ใช้วิธีการบูตสแตรปจะให้ค่าระดับความเชื่อมั่นจากการทดลองน้อยกว่ากรณีที่ใช้วิธีการบูตสแตรป ดังแสดงในภาคผนวก (Table 8-13)

ในการพิจารณาค่า MLCL ค่า MUCL และค่า MCIL จะพิจารณาว่าค่า MLCL ของวิธีการใดให้ค่ามากที่สุด ค่า MUCL และค่า MCIL ของวิธีการใดให้ค่าน้อยที่สุดจะเป็นวิธีการประมาณค่าที่เหมาะสมภายใต้สมมติฐานนั้นๆ ซึ่งผู้วิจัยสามารถสรุปได้ดัง Table ที่ 7

ข้อสรุปและข้อเสนอแนะ

จากการวิจัยครั้งนี้เราสามารถสร้างช่วงความเชื่อมั่นภายใต้สมมติฐานต่างๆ ที่เหมาะสมสำหรับค่าเฉลี่ยของประชากรที่มีการแจกแจงแบบเบ้ขวาด้วยวิธีการประมาณค่าแบบช่วงที่เหมาะสม โดยการนำข้อมูลที่มีอยู่มาคำนวณหาค่าเฉลี่ย ค่าสัมประสิทธิ์ความเบ้ ค่าเปอร์เซนต์ไทล์ที่ 100α ของตัวสถิติต่างๆ ด้วยวิธีบูตสแตรป และพิจารณาว่าเป็นไปตามเงื่อนไขการผกผันได้ของวิธีการประมาณค่าแบบช่วงด้วยตัวสถิติของจอห์นสันหรือไม่ แล้วจึงพิจารณาว่าควรใช้วิธีการประมาณค่าแบบช่วงวิธีใดในการสร้างช่วงความเชื่อมั่น การศึกษาวิจัยครั้งนี้วิธีการประมาณค่าแบบช่วงด้วยตัวสถิติของจอห์นสันจะต้องเป็นไปตามเงื่อนไขการผกผันได้ของตัวสถิติ ซึ่งในกรณีที่ไม่เป็นไปตามเงื่อนไขดังกล่าวจะไม่สามารถพิจารณาตัวสถิติของจอห์นสันได้ ดังนั้นจึงเป็นที่น่าสนใจในการพิจารณาเปรียบเทียบเฉพาะวิธีการประมาณค่าแบบช่วงด้วยตัวสถิติที วิธีการประมาณค่าแบบช่วงด้วยตัวสถิติของฮอลล์ และการประมาณค่าแบบช่วงด้วยตัวสถิติของเชน เพื่อหลีกเลี่ยงปัญหาในการผกผันได้ของตัวสถิติ

เอกสารอ้างอิง

ธีระพร วีระถาวร. 2537. ความน่าจะเป็นกับการประยุกต์. กรุงเทพมหานคร: ภาควิชาสถิติ จุฬาลงกรณ์มหาวิทยาลัย.

Andre I. Khuri. 1993. Advanced Calculus with Applications in Statistics. New York: John Wiley & Sons.

Averill, M. Law, and W. David Kelton. 1991. Simulation Modeling and Analysis. 2nd. Edition. New York: Mc Graw-Hill, Inc.

Chen, L. 1995. “Testing the mean of skewed distributions.” J. of Am Stat. Asso. 90: 767-772.

Efron, B., and Robert, J.T. 1993. An Introduction to the Bootstrap. New York: Chapman & Hall.

Hall, P. 1992. “On the removal of skewness by transformation.” J. of Ro Stat. Soc., Series B 54: 221-228.

Johnson, N.J. 1978. “Modified t Tests and Confidence Intervals for Asymmetric Populations.” Journal of American Statistical Association 73: 536-547.

Norman, L.J., Samuel, K., and N. Baralakrishnan. 1994. Continuous Univariate Distributions. New York: John Wiley & Sons.

Zhou, X.H., and Gao,S. 2000. “One-sided confidence intervals for means of positively skewed distribution.” The Am Stat. 54: 100-104.

ภาคผนวก

Table 8. The confidence level of interval estimation methods at the determined confidence level 0.95 from log-normal population. (lower confidence interval)

Skewness Coefficient	**1.0**			**3.0**			**5.0**		
Sample size	**10**	**30**	**50**	**10**	**30**	**50**	**10**	**30**	**50**
STUDENT T	0.9643 (0.868) *	0.9683 (0.947)*	0.9630 (0.972)*	0.9820 (0.805)*	0.9827 (0.990)	0.9803 (1.056)	0.9883 (0.808)*	0.9880 (1.058)	0.9867 (1.152)
JOHNSON	0.9280	0.9497	0.9500	0.9483	0.9570 (1.036)*	0.9567 (1.085)*	0.9543	0.9613 (1.129)	0.9657 (1.199)
HALL	0.9280	0.9493	0.9503	0.9477	0.9557	0.9567 (1.085)*	0.9540	0.9603 (1.130)	0.9647 (1.200)
CHEN	0.9253	0.9470	0.9497	0.9447	0.9513	0.9520	0.9507	0.9577 (1.138)*	0.9617 (1.207)*

Data in the parentheses are mean of upper confidence limit in the case that the confidence coefficient of interval estimation method is not lower than the determined confidence coefficient value.
*** is the maximum MLCL in any case.**

Table 9. The confidence level of interval estimation methods at the determined confidence level 0.95 from log-normal population. (upper confidence interval)

Skewness Coefficient	**1.0**			**3.0**			**5.0**		
Sample size	**10**	**30**	**50**	**10**	**30**	**50**	**10**	**30**	**50**
STUDENT T	0.9293	0.9390	0.9380	0.8107	0.8633	0.8760	0.7337	0.8010	0.8237
JOHNSON	0.9093	0.9380	0.9427	0.8053	0.8977	0.9087	0.7363	0.8447	0.8763
HALL	0.9087	0.9380	0.9423	0.8030	0.8940	0.9070	0.7340	0.8407	0.8727
CHEN	0.9050	0.9377	0.9423	0.7927	0.8773	0.8980	0.7190	0.8203	0.8557

Data in the parentheses are mean of upper confidence limit in the case that the confidence coefficient of interval estimation method is not lower than the determined confidence coefficient value.
*** is the maximum MUCL in any case.**

Table 10. The confidence level of interval estimation methods at the determined confidence level 0.95 from log-normal population. (two-tailed confidence interval)

Skewness Coefficient	1.0			3.0			5.0		
Sample size	**10**	**30**	**50**	**10**	**30**	**50**	**10**	**30**	**50**
STUDENT T	0.9263	0.9440	0.9480	0.8560	0.9023	0.9057	0.7817	0.8480	0.8670
JOHNSON	0.8830	0.9360	0.9387	0.8273	0.9140	0.9260	0.7677	0.8870	0.9107
HALL	0.8807	0.9357	0.9383	0.8230	0.9097	0.9230	0.7587	0.8770	0.9017
CHEN	0.8707	0.9310	0.9360	0.8020	0.8920	0.9090	0.7380	0.8497	0.8807

Data in the parentheses are mean of upper confidence limit in the case that the confidence coefficient of interval estimation method is not lower than the determined confidence coefficient value.
*** is the maximum MCIL in any case.**

Table 11. The confidence level of interval estimation methods at the determined confidence level 0.95 from weibull population. (lower confidence interval)

Skewness Coefficient	1.0			3.0			5.0		
Sample size	**10**	**30**	**50**	**10**	**30**	**50**	**10**	**30**	**50**
STUDENT T	0.9723 (0.580)*	0.9677 (0.719)*	0.9653 (0.764)*	0.9850 (0.508)*	0.9813 (0.750)*	0.9753 (0.836)	0.9920 (0.534)*	0.9847 (0.875)*	0.9777 (1.006)
JOHNSON	0.9460	0.9497	0.9550	0.9483	0.9510	0.9567 (0.883)*	0.9553	0.9563	0.9580 (1.088)
HALL	0.9447	0.9493	0.9550	0.9477	0.9510	0.9563	0.9550	0.9563	0.9567 (1.090)*
CHEN	0.9407	0.9460	0.9513	0.9450	0.9490	0.9547	0.9487	0.9513	0.9557

Data in the parentheses are mean of upper confidence limit in the case that the confidence coefficient of interval estimation method is not lower than the determined confidence coefficient value.
*** is the maximum MLCL in any case.**

Table 12. The confidence level of interval estimation methods at the determined confidence level 0.95 from weibull population. (upper confidence interval)

Skewness Coefficient	1.0			3.0			5.0		
Sample size	**10**	**30**	**50**	**10**	**30**	**50**	**10**	**30**	**50**
STUDENT T	0.8983	0.9210	0.9350	0.7640	0.8460	0.8717	0.6523	0.7393	0.7827
JOHNSON	0.8890	0.9347	0.9450	0.7750	0.8900	0.9147	0.6710	0.7923	0.8457
HALL	0.8880	0.9337	0.9443	0.7710	0.8853	0.9110	0.6623	0.7850	0.8393
CHEN	0.8820	0.9287	0.9420	0.7570	0.8683	0.8997	0.6430	0.7567	0.8133

Data in the parentheses are mean of upper confidence limit in the case that the confidence coefficient of interval estimation method is not lower than the determined confidence coefficient value.
*** is the minimum MUCL in any case.**

Table 13. The confidence level of interval estimation methods at the determined confidence level 0.95 from weibull population. (two-tailed confidence interval)

Skewness Coefficient	1.0			3.0			5.0		
Sample size	**10**	**30**	**50**	**10**	**30**	**50**	**10**	**30**	**50**
STUDENT T	0.9243	0.9443	0.9470	0.8120	0.8873	0.9033	0.7067	0.7767	0.8197
JOHNSON	0.8923	0.9370	0.9457	0.8027	0.9067	0.9277	0.7113	0.8397	0.8793
HALL	0.8897	0.9367	0.9447	0.7950	0.9007	0.9213	0.6983	0.8237	0.8663
CHEN	0.8770	0.9323	0.9423	0.7673	0.8793	0.9007	0.6657	0.7807	0.8253

Data in the parentheses are mean of upper confidence limit in the case that the confidence coefficient of interval estimation method is not lower than the determined confidence coefficient value.
*** is the minimum MCIL in any case.**