บทที่ 2

ทฤษฎีและผลงานวิจัยที่เกี่ยวข้อง

งานวิจัย เรื่อง การเปรียบเทียบวิธีการประมาณอัตราส่วนประชากร จาก การสุ่มตัวอย่างอย่างง่าย ผู้วิจัยได้ทำการศึกษาทฤษฎีและผลงานวิจัยที่เกี่ยวข้อง โดยแบ่งการนำเสนอเป็น 2 ส่วน คือส่วนที่ 1 แนวคิดและทฤษฎีที่เกี่ยวข้องกับการประมาณอัตราส่วน และส่วนที่ 2 ผลงานวิจัยที่เกี่ยวข้อง แสดงรายละเอียดดังต่อไปนี้

2.1 แนวคิดและทฤษฎีที่เกี่ยวข้องกับการประมาณอัตราส่วน

การสุ่มตัวอย่างอย่างง่าย (Simple Random Sampling)

การสุ่มตัวอย่างอย่างง่าย เป็นวิธีการเลือกตัวอย่างที่กำหนดให้ตัวอย่างในขนาดที่กำหนดทุกตัวอย่างที่อาจเป็นไปได้มีโอกาสเกิดขึ้นเท่าๆ กัน เช่น จากประชากรขนาด N ถ้าต้องการสุ่มตัวอย่างอย่างง่ายขนาด n เมื่อ n<N แบบไม่ใส่คืน จะเห็นว่ามีตัวอย่างที่อาจเป็นได้อยู่ทั้งสิ้น $\binom{N}{n}$ ตัวอย่าง แต่ละตัวอย่างที่เป็นไปได้นี้ จะได้รับโอกาสถูกเลือกเท่าๆ กัน คือด้วยความน่าจะเป็นเท่ากับ $\frac{1}{\binom{N}{n}}$ และวิธีการเลือกจะต้องก่อให้เกิดความน่าจะเป็นที่แต่ละตัวอย่าง จะถูกเลือกในลักษณะเช่นนี้ วิธีการดังกล่าวจะทำให้หน่วยแต่ละหน่วยในประชากรมีโอกาสถูกเลือกเข้ามาในตัวอย่างเท่าๆ กันด้วย (สุชาดา, 2538 น.26)

การประมาณค่า (Estimation)

การประมาณค่า คือ การคาดเดาหรือประมาณค่าพารามิเตอร์ในประชากรที่เราไม่ทราบค่า โดยอาศัยความรู้เกี่ยวกับตัวสถิติและการแจกแจงของตัวสถิติ การประมาณค่าพารามิเตอร์ เราสามารถประมาณได้ 2 แบบ คือ การประมาณค่าแบบจุดและการประมาณค่าแบบช่วง

การประมาณค่าแบบจุด (Point estimation) เป็นการประมาณค่าพารามิเตอร์โดยใช้ตัวเลขเพียงค่าเดียว ซึ่งได้จากการคำนวณค่าสถิติหนึ่งจากตัวอย่าง ที่คาดว่าจะมีค่าใกล้กับ

ค่าพารามิเตอร์ เพื่อใช้ตัวสถิตินี้เป็นตัวประมาณค่าพารามิเตอร์ เรียกตัวสถิตินี้ว่า “ตัวประมาณ” (Estimator) ตัวประมาณ ก็คือ ตัวสถิติที่ใช้ประมาณค่าพารามิเตอร์ในประชากร ส่วนค่าของตัวสถิติที่ได้จากการคำนวณจากตัวอย่าง เรียกว่า “ค่าประมาณ” (Estimate)

การประมาณแบบช่วง (Interval estimation) คือ การประมาณโดยใช้ช่วงสุ่มที่คาดว่าจะครอบคลุมค่าพารามิเตอร์ที่เราต้องการประมาณ โดยช่วงสุ่มนั้นมีโอกาสครอบคลุมค่าพารามิเตอร์ด้วยความน่าจะเป็นที่กำหนดไว้ล่วงหน้า ความน่าจะเป็นนี้เรียกว่า “ระดับความเชื่อมั่น” (Level of Confidence) ซึ่งมักกำหนดให้มีค่าสูง เช่น 0.90, 0.95 ,หรือ 0.99 เป็นต้น ถ้าให้ θ เป็นพารามิเตอร์ที่เราต้องการประมาณค่า ให้ L และ U เป็นตัวแปรสุ่ม โดย L มีค่าน้อยกว่า θ และ U มีค่ามากกว่า θ เรากำหนดให้ช่วงสุ่ม (L, U) นี้มีโอกาสที่ครอบคลุมค่า θ ด้วยความน่าจะเป็น $1-\alpha$ โดยที่ $0<\alpha<1$ ซึ่งสามารถเขียนสัญลักษณ์แทนได้ด้วย $P(L<\theta<U)=1-\alpha$ เราให้ช่วงสุ่ม (L, U) หมายถึง ช่วงความเชื่อมั่น (Confidence Interval) (1-α)100% ของพารามิเตอร์ θ และเรียก $(1-\alpha)$ ว่า “ระดับความเชื่อมั่น” (Level of Confidence) ทั้งนี้ค่าของ L และ U จะขึ้นอยู่กับการแจกแจงของตัวสถิติที่เป็นตัวประมาณแบบจุดของ θ ช่วงความเชื่อมั่น $(1-\alpha)$100% ของพารามิเตอร์ θ นี้หมายถึง ในการทดลองสุ่มตัวอย่างเพื่อสร้างช่วงสุ่ม 100 ครั้ง จะมีช่วงสุ่ม $(1-\alpha)$100 ช่วงสุ่ม ที่ครอบคลุมค่าพารามิเตอร์ θ และมีช่วงสุ่มจำนวน (α)100 ช่วงสุ่มที่ไม่ครอบคลุมค่าพารามิเตอร์ θ เช่น ช่วงความเชื่อมั่น 95% หมายความว่า ถ้าทำการทดลองสุ่มตัวอย่างเพื่อสร้างช่วงสุ่ม 1000 ครั้ง จะมี 950 ช่วงสุ่มที่ครอบคลุม ค่าพารามิเตอร์ และอีก 50 ช่วงสุ่ม จะไม่ครอบคลุมค่าพารามิเตอร์ (กัลยา, 2549 น.243-245)

พารามิเตอร์ (Parameter) คือ ค่าคงที่ที่แสดงคุณลักษณะบางประการของประชากร เช่น μ , $\pi(P)$ หรือ R ในที่นี้ใช้ อัตราส่วนของประชากร (Population Ratio) คือ

$$R=\frac{\mu_y}{\mu_x}$$

ตัวสถิติ (Statistic) คือ ฟังก์ชันของค่าสังเกตที่วัดมาจากหน่วยตัวอย่างต่างๆ ที่ถูกเลือกขึ้นมาเป็นตัวอย่าง ฟังก์ชันดังกล่าวจะไม่มีตัวพารามิเตอร์อื่นใดที่ยังไม่ทราบค่าติดอยู่เลย ตัวสถิติที่ได้จะใช้เป็นตัวประมาณ (Estimator) ของพารามิเตอร์ ค่าของตัวสถิติที่คำนวณออกมาเป็นตัวเลขจะใช้เป็นค่าประมาณ (Estimate) ของพารามิเตอร์ที่ยังไม่ทราบค่าต่อไป เช่น $\bar{x}$, $\hat{p}$ หรือ r ในที่นี้ใช้ ตัวประมาณอัตราส่วนของประชากร คือ $r_1=\hat{R}=\frac{\bar{Y}}{\bar{X}}$ และ $r_2=\hat{R}=\frac{1}{n}\sum_{i=1}^{n}\frac{Y_i}{X_i}$

ตัวประมาณอัตราส่วนประชากร

การประมาณอัตราส่วนประชากร อาจใช้ตัวประมาณ $r_1 = \hat{R} = \frac{\bar{Y}}{\bar{X}}$ โดยที่ r_1 เป็นตัวประมาณที่ใช้กันอย่างแพร่หลาย ถึงแม้ว่าจะมีความเอนเอียงในขนาดของตัวอย่างที่มี ขนาดเล็ก แต่ความเอนเอียงจะลดลง เมื่อขนาดของตัวอย่างมีขนาดใหญ่

โดยความแปรปรวนประชากรหาได้จาก $\text{var}(r_1) \cong \frac{1-\frac{n}{N}}{n\mu_x^2} \sum_{i=1}^{N} \frac{\left(y_i - \frac{\mu_y}{\mu_x} x_i\right)^2}{N-1}$

ความแปรปรวนตัวอย่างสามารถหาได้จาก $\widehat{\text{var}}(r_1) = \frac{1-\frac{n}{N}}{n\mu_x^2} \sum_{i=1}^{n} \frac{(y_i - r_1 x_i)^2}{N-1}$

แต่ถ้าไม่ทราบค่าเฉลี่ยประชากร ก็ใช้ค่าเฉลี่ยตัวอย่างแทน

ความแปรปรวนตัวอย่างปรับค่า คือ $\widetilde{\text{var}}(r_1) = \frac{1-\frac{n}{N}}{n\bar{X}^2} \sum_{i=1}^{n} \frac{(y_i - r_1 x_i)^2}{n-1}$ (Thompson,1992 น.61)

ทฤษฎีบท ในการสุ่มตัวอย่างอย่างง่าย ขนาด n เมื่อ n มีขนาดใหญ่จากประชากรขนาด N และวัดค่า (x_i, y_i) จากหน่วยตัวอย่างทุกๆ หน่วยเพื่อประมาณค่า R ด้วย $r_1 = \hat{R} = \frac{\bar{Y}}{\bar{X}}$ แล้วค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) และค่าความแปรปรวนของ $r_1 = \hat{R} = \frac{\bar{Y}}{\bar{X}}$ โดยประมาณ

คือ $\text{MSE}(r_1) \cong \text{var}(r_1) \cong \frac{1-\frac{n}{N}}{n\mu_x^2} \sum_{i=1}^{N} \frac{\left(y_i - \frac{\mu_y}{\mu_x} x_i\right)^2}{N-1}$

<u>พิสูจน์</u> $\because \quad \hat{R} - R = \frac{\bar{y}}{\bar{x}} - R = \frac{\bar{y} - R\bar{x}}{\bar{x}}$ เมื่อ $R = \frac{\mu_y}{\mu_x}$

ถ้า n มีขนาดใหญ่พอ $\bar{x}$ ควรจะมีค่าใกล้เคียงกับ μ_x จนทำให้สามารถแทนค่า $\bar{x}$ ด้วย μ_x ซึ่งเป็นค่าคงที่ได้ ทำให้ $\hat{R} - R \cong \frac{1}{\mu_x}(\bar{y} - R\bar{x})$ ภายใต้เงื่อนไขนี้ จะได้ว่า

$$\begin{aligned} E(\hat{R} - R) &\cong E\left(\frac{1}{\mu_x}(\bar{y} - R\bar{x})\right) \\ &= \frac{1}{\mu_x} E(\bar{y} - R\bar{x}) \\ &= \frac{1}{\mu_x}\left[E(\bar{y}) - E(R\bar{x})\right] \\ &= \frac{1}{\mu_x}\left[E(\bar{y}) - RE(\bar{x})\right] \end{aligned}$$

$$= \frac{1}{\mu_x}\left[\mu_y - R\mu_x\right]$$
$$= \frac{1}{\mu_x}\left[\mu_y - \mu_y\right]$$
$$= 0$$

จึงสรุปได้ว่า ถ้า n มีขนาดใหญ่พอ ความเอนเอียงของตัวประมาณ $r_1 = \hat{R} = \frac{\bar{Y}}{\bar{X}}$ จะมีขนาดเล็กจนไม่ต้องพิจารณาได้

เนื่องจาก $MSE\left(\hat{R}_1\right) = E\left(\hat{R}_1 - R\right)^2$

$$\cong \frac{1}{\mu_x^2} E\left(\bar{y} - R\bar{x}\right)^2$$

ถ้ากำหนดให้ $d_i = y_i - Rx_i$

จะได้ $\bar{d} = \bar{y} - R\bar{x}$

และ $\bar{D} = \mu_y - R\mu_x = 0$

ดังนั้น $E\left(\bar{y} - R\bar{x}\right)^2 = E\left(\bar{d} - \bar{D}\right)^2 = var\left(\bar{d}\right)$

และ $\bar{d}$ ย่อมมีคุณสมบัติเช่นเดียวกับ $\bar{y}$ หากพิจารณาว่า d_i คือค่าที่เก็บได้จากหน่วย i ในตัวอย่างสุ่มอย่างอย่างง่ายขนาด n

$\therefore$ $var\left(\bar{d}\right) = \left(1 - \frac{n}{N}\right)\frac{\sigma_d^2}{n}$ เมื่อ $\sigma_d^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(d_i - \bar{D}\right)^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(y_i - Rx_i\right)^2$

ดังนั้น $MSE\left(r_1\right) \cong var\left(r_1\right) \cong \frac{1 - \frac{n}{N}}{n\mu_x^2}\sum_{i=1}^{N}\frac{\left(y_i - \frac{\mu_y}{\mu_x}x_i\right)^2}{N-1}$ (สุชาดา,2538 น. 52-54)

สำหรับตัวอย่าง ที่มีขนาดเล็ก การประมาณช่วงความเชื่อมั่น $100(1-\alpha)\%$ สำหรับอัตราส่วนประชากร (R) คือ $r_1 \pm t_{n-1,\left(\frac{\alpha}{2}\right)}\sqrt{\widetilde{var}\left(r_1\right)}$ โดยที่ $t_{n-1,\left(\frac{\alpha}{2}\right)}$ เปิดตารางการแจกแจง Student t ที่ $\left(\frac{\alpha}{2}\right)$ โดยมีองศาอิสระ (degrees of freedom) เป็น $n-1$ สำหรับตัวอย่าง ที่มีขนาดใหญ่ การประมาณช่วงความเชื่อมั่น $100(1-\alpha)\%$ สำหรับค่าเฉลี่ยประชากร (R) คือ $r_1 \pm z_{\left(\frac{\alpha}{2}\right)}\sqrt{\widetilde{var}\left(r_1\right)}$

และการประมาณอัตราส่วนประชากร อาจใช้ตัวประมาณ $r_2 = \hat{R} = \frac{1}{n}\sum_{i=1}^{n}\frac{Y_i}{X_i}$ โดยที่ r_2 เป็นตัวประมาณที่ไม่แพร่หลาย

โดยมีความแปรปรวนของประชากร คือ $\text{var}(r_2)=\left(1-\frac{n}{N}\right)\frac{\sum_{i=1}^{N}\left(R_i-\bar{R}\right)^2}{n(N-1)}$

เมื่อ $R_i=\frac{Y_i}{X_i}$ และ $\bar{R}=\frac{\sum_{i=1}^{N}R_i}{N}$

ความแปรปรวนของตัวอย่าง คือ $\widehat{\text{var}}(r_2)=\left(1-\frac{n}{N}\right)\frac{\sum_{i=1}^{n}\left(r_i-\bar{r}\right)^2}{n(n-1)}=\left(1-\frac{n}{N}\right)\frac{\sum_{i=1}^{n}\left(r_i-r_2\right)^2}{n(n-1)}$

เมื่อ $r_i=\frac{Y_i}{X_i}$ และ $\bar{r}=\frac{\sum_{i=1}^{n}r_i}{n}=\frac{1}{n}\sum_{i=1}^{n}\frac{Y_i}{X_i}=r_2$

จาก Christian Salas. และ Timothy G. Gregoire ได้วิจัยเรื่อง Statistical analysis of ratio estimators and their estimators of variances when the auxiliary variate is measured with error โดยใช้ $\text{var}(\hat{\tau}_{y2})=\text{var}(r_2\tau_x)=\left(\frac{1}{n}-\frac{1}{N}\right)\tau_x^2\left[\frac{1}{N-1}\sum_{i=1}^{N}\left(R_i-\bar{R}\right)^2\right]$ และก็ยังได้ใช้ $\widehat{\text{var}}(\hat{\tau}_{y2})=\widehat{\text{var}}(r_2\tau_x)=\left(\frac{1}{n}-\frac{1}{N}\right)\tau_x^2\frac{1}{n-1}\sum_{i=1}^{n}(r_i-\bar{r})^2$ เมื่อ τ_x แทนค่าผลรวมประชากรของตัวแปรช่วย ดังนั้นการศึกษาครั้งนี้ได้ใช้ความแปรปรวนที่กล่าวมา นำไปใช้ในการหาค่าของ $\text{var}(r_2)$ และ $\widehat{\text{var}}(r_2)$ ตามลำดับ โดยที่สามารถแสดงได้ดังนี้

การหาค่าของ $\text{var}(r_2)$

จาก
$$\text{var}(r_2\tau_x)=\left(\frac{1}{n}-\frac{1}{N}\right)\tau_x^2\left[\frac{1}{N-1}\sum_{i=1}^{N}\left(R_i-\bar{R}\right)^2\right]$$
$$\tau_x^2\,\text{var}(r_2)=\left(\frac{1}{n}-\frac{1}{N}\right)\tau_x^2\left[\frac{1}{N-1}\sum_{i=1}^{N}\left(R_i-\bar{R}\right)^2\right]$$
$$\text{var}(r_2)=\frac{1}{\tau_x^2}\left(\frac{1}{n}-\frac{1}{N}\right)\tau_x^2\left[\frac{1}{N-1}\sum_{i=1}^{N}\left(R_i-\bar{R}\right)^2\right]$$
$$=\left(\frac{1}{n}-\frac{1}{N}\right)\left[\frac{1}{N-1}\sum_{i=1}^{N}\left(R_i-\bar{R}\right)^2\right]$$
$$=\left(\frac{N-n}{nN}\right)\left[\frac{1}{N-1}\sum_{i=1}^{N}\left(R_i-\bar{R}\right)^2\right]$$
$$=\left(\frac{N-n}{nN}\right)\left[\frac{1}{N-1}\sum_{i=1}^{N}\left(R_i-\bar{R}\right)^2\right]$$
$$=\left(1-\frac{n}{N}\right)\left[\frac{1}{n(N-1)}\sum_{i=1}^{N}\left(R_i-\bar{R}\right)^2\right]$$

ดังนั้น $\text{var}(r_2) = \left(1-\frac{n}{N}\right)\frac{\sum_{i=1}^{N}\left(R_i-\bar{R}\right)^2}{n(N-1)}$

และ การหาค่าของ $\widehat{\text{var}}(r_2)$

จาก $\widehat{\text{var}}(r_2\tau_x) = \left(\frac{1}{n}-\frac{1}{N}\right)\tau_x^2\frac{1}{n-1}\sum_{i=1}^{n}(r_i-\bar{r})^2$

$$\tau_x^2\widehat{\text{var}}(r_2) = \left(\frac{1}{n}-\frac{1}{N}\right)\tau_x^2\frac{1}{n-1}\sum_{i=1}^{n}(r_i-\bar{r})^2$$

$$\widehat{\text{var}}(r_2) = \frac{1}{\tau_x^2}\left(\frac{1}{n}-\frac{1}{N}\right)\tau_x^2\frac{1}{n-1}\sum_{i=1}^{n}(r_i-\bar{r})^2$$

$$= \left(\frac{1}{n}-\frac{1}{N}\right)\frac{1}{n-1}\sum_{i=1}^{n}(r_i-\bar{r})^2$$

$$= \left(\frac{N-n}{nN}\right)\frac{1}{n-1}\sum_{i=1}^{n}(r_i-\bar{r})^2$$

$$= \frac{1}{n}\left(\frac{N-n}{N}\right)\frac{1}{n-1}\sum_{i=1}^{n}(r_i-\bar{r})^2$$

$$= \frac{1}{n}\left(1-\frac{n}{N}\right)\frac{1}{n-1}\sum_{i=1}^{n}(r_i-\bar{r})^2$$

$$= \left(1-\frac{n}{N}\right)\frac{1}{n(n-1)}\sum_{i=1}^{n}(r_i-\bar{r})^2$$

ดังนั้น $\widehat{\text{var}}(r_2) = \left(1-\frac{n}{N}\right)\frac{\sum_{i=1}^{n}\left(r_i-\bar{r}\right)^2}{n(n-1)} = \left(1-\frac{n}{N}\right)\frac{\sum_{i=1}^{n}\left(r_i-r_2\right)^2}{n(n-1)}$

(Christian Salas. และ Timothy G. Gregoire,2009)

สำหรับตัวอย่าง ที่มีขนาดเล็ก การประมาณช่วงความเชื่อมั่น $100(1-\alpha)\%$ สำหรับอัตราส่วนประชากร (R) คือ $r_2 \pm t_{n-1,\left(\frac{\alpha}{2}\right)}\sqrt{\widehat{\text{var}}\left(r_2\right)}$ โดยที่ $t_{n-1,\left(\frac{\alpha}{2}\right)}$ เปิดตารางการแจกแจง Student t ที่ $\left(\frac{\alpha}{2}\right)$ โดยมีองศาอิสระ (degrees of freedom) เป็น $n-1$ สำหรับตัวอย่าง ที่มีขนาดใหญ่ การประมาณช่วงความเชื่อมั่น $100(1-\alpha)\%$ สำหรับค่าเฉลี่ยประชากร (R) คือ $r_2 \pm z_{\left(\frac{\alpha}{2}\right)}\sqrt{\widehat{\text{var}}\left(r_2\right)}$

2.2 การแจกแจงที่ใช้ในการศึกษา

การแจกแจงปกติสองตัวแปร (Bivariate Normal Distribution)

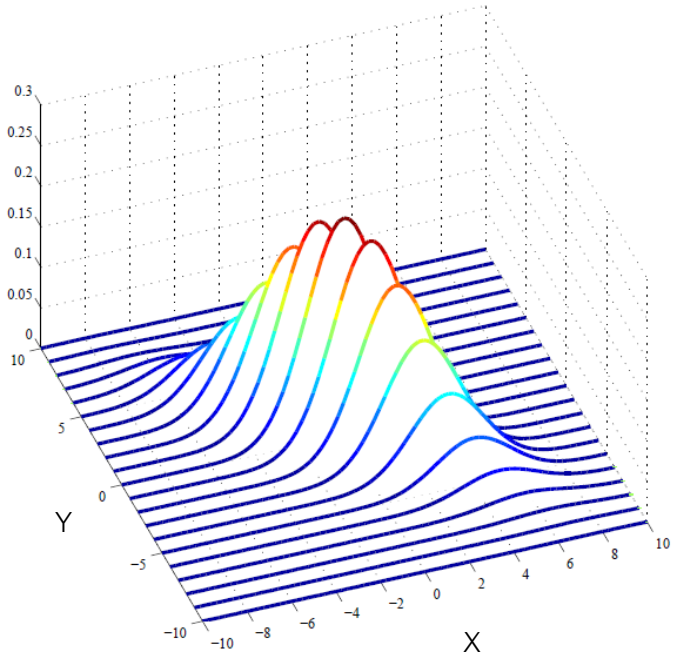
กำหนดให้ X เป็นตัวแปรสุ่มต่อเนื่องที่มีการแจกแจงปกติ โดยมีพารามิเตอร์ μ_x และ σ_x^2 และ Y เป็นตัวแปรสุ่มต่อเนื่องที่มีการแจกแจงปกติ โดยมีพารามิเตอร์ μ_y และ σ_y^2 ดังนั้น ฟังก์ชันการแจกแจง (Distribution function) หรือ ฟังก์ชันความหนาแน่นความน่าจะเป็น (Probability density function) ของ X และ Y คือ

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2+\left(\frac{y-\mu_y}{\sigma_y}\right)^2-2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right)\right]},$$

$$-\infty \le x, y < \infty \quad , \quad -\infty < \mu_x, \mu_y < \infty \quad , \quad \sigma_x^2, \sigma_y^2 > 0$$

คุณสมบัติของการแจกแจงปกติ มีดังนี้

1. ฟังก์ชันความหนาแน่นความน่าจะเป็นมีลักษณะเป็นเส้นโค้งไม่มีความเบ้ หรือมีค่าความเบ้เท่ากับศูนย์
2. เส้นโค้งปกติเป็นรูประฆังคว่ำสมมาตรรอบค่าเฉลี่ย
3. ค่าเฉลี่ย มัธยฐาน และฐานนิยมมีค่าเท่ากันและอยู่ในตำแหน่งกลางเส้นโค้งในแนวตั้งฉาก
4. พื้นที่ใต้โค้งทั้งหมดคือความน่าจะเป็นของ ปริภูมิตัวอย่าง (Sample space) ซึ่งมีค่าเท่ากับ 1 หรือ 100%
5. พื้นที่ใต้โค้งระหว่างจุดที่เป็น $\pm 1\sigma$ มีประมาณ 68.26%, พื้นที่ใต้โค้งระหว่างจุดที่เป็น $\pm 2\sigma$ มีประมาณ 95.46% และพื้นที่ใต้โค้งระหว่างจุดที่เป็น $\pm 3\sigma$ มีประมาณ 99.74%
6. เมื่อการแจกแจงปกติมีค่า $\mu = 0$ และ $\sigma^2 = 1$ จะมีการแจกแจงที่เรียกว่าการแจกแจงปกติมาตรฐาน (Standard normal distribution)



ภาพที่ 2.1 แสดงฟังก์ชันความหนาแน่นของการแจกแจงปกติสองตัวแปร

การแจกแจงปัวซง (Poisson Distribution)

กำหนดให้ X เป็นตัวแปรสุ่ม ไม่ต่อเนื่องที่มีการแจกแจงปัวซง โดยมีพารามิเตอร์ λ ดังนั้น ฟังก์ ์ การแจกแจง (Distribution function) หรือ ฟังก์ชันความหนาแน่นความน่าจะเป็น (Probability density function) ของ X คือ

$$f(x;\lambda) = \frac{\lambda^{x} e^{-\lambda}}{x!} \quad , \quad x = 0,1,2,3,... \quad ,\lambda > 0$$

ลักษณะทั่วไปของตัวแปรสุ่ม X ที่มีการแจกแจงปัวซง

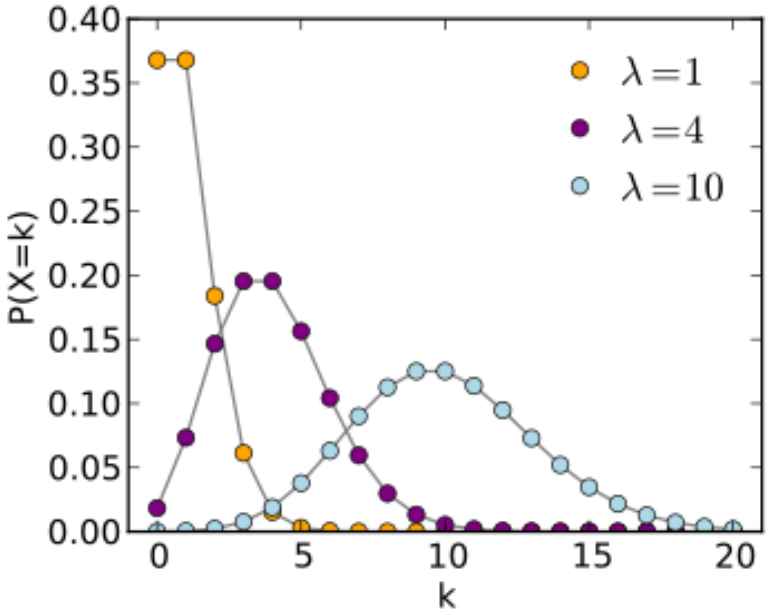
- ค่าเฉลี่ยของตัวแปรสุ่ม X

  $E(X) = \lambda$
- ความแปรปรวนของตัวแปรสุ่ม X

  $Var(X) = \lambda$
- สัมประสิทธิ์ความเบ้

  $\gamma_1 = \lambda^{-\frac{1}{2}}$
- สัมประสิทธิ์ความโด่ง

  $\gamma_2 = \lambda^{-1}$



ภาพที่ 2.2 แสดงฟังก์ชันความหนาแน่นของการแจกแจงปัวซง

การแจกแจงล็อกนอร์มัล (Lognormal Distribution)

กำหนดให้ X เป็นตัวแปรสุ่มต่อเนื่องที่มีการแจกแจงล็อกนอร์มั ล โดยมีพารามิเตอร์ μ และ σ^2 ดังนั้น ฟังก์ชันการแจกแจง (Distribution function) หรือ ฟังก์ชันความหนาแน่นความน่าจะเป็น (Probability density function) ของ X คือ

$$f(x;\mu,\sigma^2) = \frac{1}{x\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{\ln x-\mu}{\sigma}\right)^2} \quad , \quad 0 \le x < \infty \, , \, -\infty < \mu < \infty \, , \, \sigma^2 > 0$$

ลักษณะทั่วไปของตัวแปรสุ่ม X ที่มีการแจกแจงล็อกนอร์มัล

- ค่าเฉลี่ยของตัวแปรสุ่ม X

  $E(X) = \exp\left(\mu + \frac{\sigma^2}{2}\right)$

- ความแปรปรวนของตัวแปรสุ่ม X

  $Var(X) = m^2 \omega \ (\omega\text{-}1)$

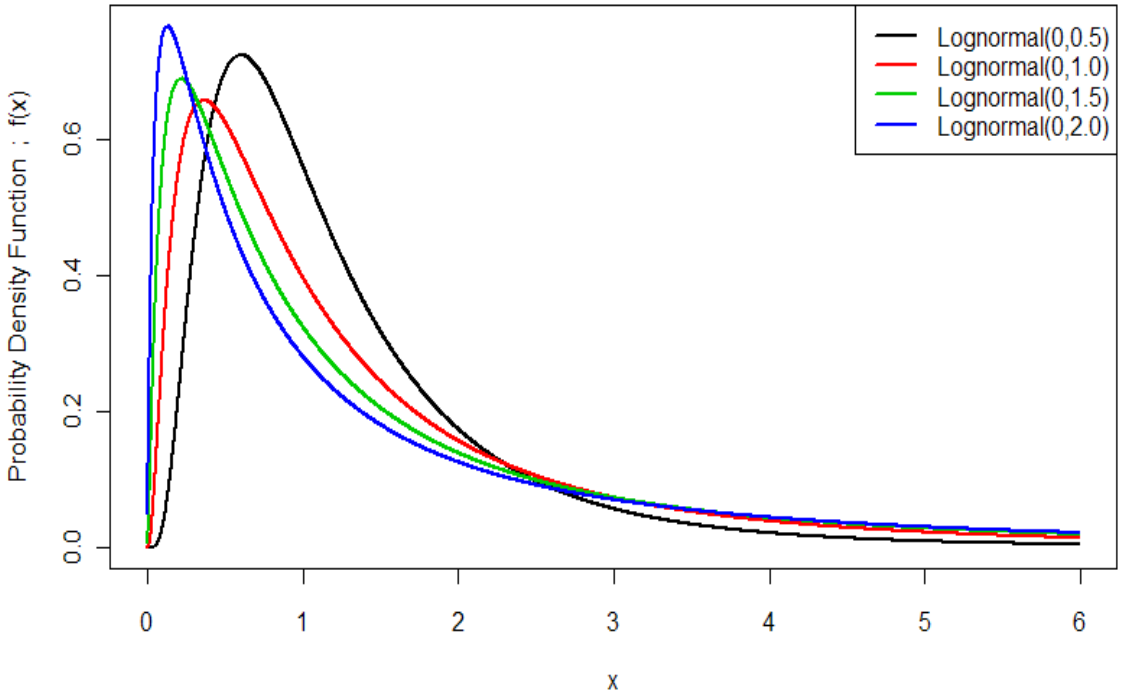
- สัมประสิทธิ์ความเบ้

  $\gamma_1 = (\omega+2)\sqrt{\omega-1}$

- สัมประสิทธิ์ความโด่ง

$$\gamma_2 = \omega^4 + 2\omega^3 + 3\omega^2 - 3$$

โดยที่ $m = \exp(\mu)$, $\omega = \exp(\sigma^2)$



ภาพที่ 2.3 แสดงฟังก์ชันความหนาแน่นของการแจกแจงล็อกนอร์มัล

การแจกแจงโคชี (Cauchy Distribution)

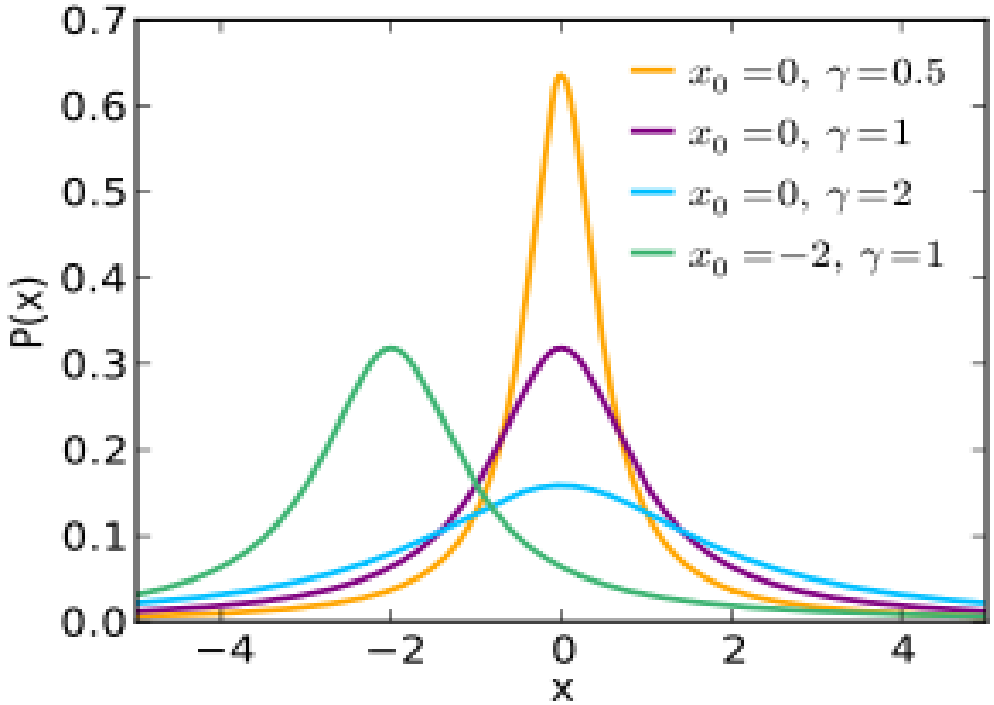
กำหนดให้ X เป็นตัวแปรสุ่มต่อเนื่องที่มีการแจกแจงโคชีโดยมีพารามิเตอร์ χ_0 และ γ ดังนั้น ฟังก์ชันการแจกแจง (Distribution function) หรือ ฟังก์ชันความหนาแน่นความน่าจะเป็น (Probability density function) ของ X คือ

$$f(x;\chi_0,\gamma) = \frac{1}{\pi\gamma\left[1+\left(\frac{x-\chi_0}{\gamma}\right)^2\right]} \quad , \quad 0 \le x < \infty \, , \, -\infty < \chi_0 < \infty \, , \, \gamma > 0$$

ลักษณะทั่วไปของตัวแปรสุ่ม X ที่มีการแจกแจงโคชี

- ค่าเฉลี่ยของตัวแปรสุ่ม x ไม่นิยาม

- ความแปรปรวนของตัวแปรสุ่ม x ไม่นิยาม
- สัมประสิทธิ์ความเบ้ ไม่นิยาม
- สัมประสิทธิ์ความโด่ง ไม่นิยาม



ภาพที่ 2.4 แสดงฟังก์ชันความหนาแน่นของการแจกแจงโคชี

2.3 ผลงานวิจัยที่เกี่ยวข้อง

งานวิจัยที่เกี่ยวข้องกับการใช้ค่าเฉลี่ยตัวอย่างของตัวแปรช่วยแทนค่าเฉลี่ยประชากรของตัวแปรช่วยยังไม่มีผู้ใดได้ศึกษา แต่งานวิจัยที่เกี่ยวข้องกับเปรียบเทียบการหาช่วงความเชื่อมั่นและการประมาณอัตราส่วนประชากรมีดังนี้

งานวิจัยที่เกี่ยวข้องกับเปรียบเทียบการหาช่วงความเชื่อมั่น

สุกัญญา อินทรภักดิ์ และ พิษณุ เจียวคุณ(2548) ได้ศึกษาและเปรียบเทียบวิธีการประมาณค่าช่วงความเชื่อมั่นสำหรับพารามิเตอร์ของข้อมูลชั่วชีวิต ซึ่งมีการแจกแจงล็อกนอร์มัลและการแจกแจงไวบูลล์โดยใช้ความกว้างเฉลี่ยของช่วงความเชื่อมั่นเป็นเกณฑ์ในการเปรียบเทียบวิธีการประมาณค่าช่วงความเชื่อมั่น วิธีการประมาณที่ใช้ในการศึกษาครั้งนี้ คือ วิธี Asymptotic Normality (AN) วิธี Uncorrected Likelihood Ratio (ULR) และวิธี Signed

Square Roots of the Likelihood Ratio (SLR) โดยศึกษาข้อมูลสมบูรณ์และข้อมูลเซ็นเซอร์ประเภทที่ 2 10% , 20% และ 50% ที่ขนาดตัวอย่างเท่ากับ 10, 20, 30 และ 40 กำหนดระดับความเชื่อมั่นเท่ากับ 95% และกำหนดค่าพารามิเตอร์ μ, σ, α และ β เท่ากับ 7, 3, 100 และ 5 ตามลำดับ กระทำการทดลองซ้ำ 1,000 ครั้งในแต่ละสถานการณ์ ซึ่งผลการศึกษาสรุปได้ดังนี้

กรณีข้อมูลสมบูรณ์ สำหรับทุกค่าพารามิเตอร์ที่ระดับความเชื่อมั่น 95% พบว่า เมื่อขนาดตัวอย่างเท่ากับ 10 และ 20 วิธี Uncorrected Likelihood Ratio ให้ความกว้างเฉลี่ยของช่วงความเชื่อมั่นแคบที่สุด ขนาดตัวอย่างเท่ากับ 30 และ 40 วิธี Asymptotic Normality ให้ความกว้างเฉลี่ยของช่วงความเชื่อมั่นแคบที่สุด

กรณีข้อมูลเซ็นเซอร์ประเภทที่ 2 สำหรับทุกค่าพารามิเตอร์ และทุกขนาดตัวอย่างที่ระดับความเชื่อมั่น 95% พบว่า เมื่อข้อมูลเซ็นเซอร์ 10% วิธี Signed Square Roots of the Likelihood Ratio ให้ความกว้างเฉลี่ยของช่วงความเชื่อมั่นแคบที่สุด

สำหรับทุกค่าพารามิเตอร์ ทุกวิธีการประมาณ และทุกขนาดตัวอย่าง พบว่า ในข้อมูลสมบูรณ์และข้อมูลเซ็นเซอร์ประเภทที่ 2 10% จะให้ร้อยละที่ช่วงความเชื่อมั่นคลุมค่าพารามิเตอร์อยู่ระหว่างช่วงความเชื่อมั่นของร้อยละที่ช่วงความเชื่อมั่นคลุมค่าพารามิเตอร์ที่กำหนด ข้อมูลเซ็นเซอร์ประเภทที่ 2 20% และ 50% จะให้ร้อยละที่ช่วงความเชื่อมั่นคลุมค่าพารามิเตอร์อยู่นอกช่วงความเชื่อมั่นของร้อยละที่ช่วงความเชื่อมั่นคลุมค่าพารามิเตอร์ที่กำหนด โดยที่ร้อยละที่ช่วงความเชื่อมั่นคลุมค่าพารามิเตอร์จะมีแนวโน้มลดลง เมื่อระดับการเซ็นเซอร์เพิ่มขึ้น

วราภรณ์ กรงทอง (2551) ได้เปรียบเทียบวิธีการประมาณช่วงความเชื่อมั่นสำหรับอัตราส่วนออดส์ในการถดถอยโลจิสติกพหุคูณ โดยใช้วิธีการประมาณค่า 3 วิธี คือ วิธีแบบฉบับ (Classical Method) วิธีปริมาณหมุน (Pivotal Quantity Method) ที่ได้รับการปรับโดย Wilson และ Langenberg และวิธีเบย์เซียน (Bayesian Method) เกณฑ์ในการเปรียบเทียบ คือ ค่าสัมประสิทธิ์ความเชื่อมั่นและความกว้างเฉลี่ยของช่วงความเชื่อมั่นอัตราส่วนออดส์ มีตัวแปรอิสระ 2 ตัว x_1 มีการแจกแจงแบบแบร์นูลลีและ x_2 มีการแจกแจงแบบเลขชี้กำลัง ค่าพารามิเตอร์ของ β_0 เท่ากับ $-(\beta_1+\beta_2)/2$, β_1 และ β_2 เท่ากับ 0.3, 0.9 และ 1.5 ขนาดตัวอย่าง (n) เท่ากับ 20, 30, 40,50, 60, 70, 80, 90, 100, 110, 120, 130, 140 และ 150 กำหนดสัมประสิทธิ์ความเชื่อมั่นเท่ากับ 0.90,0.95 และ 0.99 การเปรียบเทียบกระทำภายใต้เงื่อนไขการเปลี่ยนแปลงของค่าพารามิเตอร์ ขนาดตัวอย่าง และสัมประสิทธิ์ความเชื่อมั่น ข้อมูลที่ใช้ในการวิจัยครั้งนี้ได้จากการจำลองโดยเทคนิคมอนติคาร์โลและกระทำซ้ำ 1,000 ครั้งในแต่ละสถานการณ์ผลการวิจัยสรุปได้ดังนี้

1) เมื่อพิจารณาสัมประสิทธิ์ความเชื่อมั่น

พบว่าการประมาณด้วยวิธีแบบฉบับ วิธีปริมาณหมุนที่ได้รับการปรับโดย Wilson และLangenberg และวิธีเบย์เซียน ให้ค่าสัมประสิทธิ์ความเชื่อมั่นของพารามิเตอร์อัตราส่วนออดส์ในตัวแปรอิสระทั้ง 2 ตัว ไม่ต่ำกว่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนดในทุก ๆ สถานการณ์ เมื่อค่าของพารามิเตอร์เพิ่มขึ้น ทำให้ค่าสัมประสิทธิ์ความเชื่อมั่นเพิ่มขึ้น

2) เมื่อพิจารณาความกว้างเฉลี่ยของช่วงความเชื่อมั่น

สำหรับทุกค่าพารามิเตอร์ ทุกขนาดตัวอย่างและทุกสัมประสิทธิ์ความเชื่อมั่น พบว่า การประมาณด้วยวิธีปริมาณหมุนที่ได้รับการปรับโดย Wilson และ Langenberg ให้ความกว้างเฉลี่ยของช่วงความเชื่อมั่นสำหรับอัตราส่วนออดส์ของตัวแปรอิสระทั้ง 2 ตัว แคบที่สุดเมื่อตัวอย่างขนาดใหญ่ ที่สัมประสิทธิ์ความเชื่อมั่น 0.90 และ 0.95 ($n \geq 80$) และสัมประสิทธิ์ความเชื่อมั่น 0.99 ($n \geq 70$) ทุกวิธีการประมาณให้ค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่นสำหรับอัตราส่วนออดส์ของตัวแปรอิสระทั้ง 2 ตัว ใกล้เคียงกันในทุก ๆ สถานการณ์เมื่อขนาดตัวอย่างเพิ่มขึ้น ทำให้ค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่นสำหรับอัตราส่วนออดส์ทั้ง 2 ตัว ลดลง แสดงว่าความกว้างเฉลี่ยของช่วงความเชื่อมั่นสำหรับอัตราส่วนออดส์แปรผกผันกับขนาดตัวอย่าง และเมื่อสัมประสิทธิ์ความเชื่อมั่นเพิ่มขึ้นทำให้ความกว้างเฉลี่ยของช่วงความเชื่อมั่นเพิ่มขึ้น แสดงว่าค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่นแปรผันตามสัมประสิทธิ์ความเชื่อมั่น

วราฤทธิ์ พานิชกิจโกศลกุล (2550) ได้ศึกษาการเปรียบเทียบวิธีการประมาณค่าช่วงความเชื่อมั่นของพารามิเตอร์ของการแจกแจงแบบเลขชี้กำลัง เมื่อข้อมูลมีค่าผิดปกติ ด้วยวิธีการประมาณ 3 วิธี คือ วิธีการประมาณแบบปกติ (Normal method) วิธีการประมาณด้วยรากของสมการกำลังสอง (Root of Quadratic Equation method) และวิธีการประมาณด้วยช่วงแบบเบส์ (Bayesian Interval method) โดยการเปรียบเทียบค่าประมาณสัมประสิทธิ์ความเชื่อมั่นและค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่น เมื่อขนาดตัวอย่างเท่ากับ 20, 40, 100 และ 20 ค่าพารามิเตอร์ θ มีค่าตั้งแต่ 1 ถึง 10 โดยเพิ่มขึ้นครั้งละ 1 ค่าผิดปกติมาจากข้อมูลที่มีการแจกแจงไคกำลังสองและล็อกนอร์มัล กำหนดร้อยละของค่าผิดปกติเท่ากับ 5 และ 10 และคำนวณช่วงความเชื่อมั่นที่ระดับความเชื่อมั่นร้อยละ 95 การวิจัยครั้งนี้ใช้วิธีการจำลองแบบมอนติคาร์โลและทำการทดลองซ้ำๆ กัน 1,000 ครั้ง ในแต่ละสถานการณ์ ผลการวิจัยสรุปได้ดังนี้

วิธีการประมาณแบบปกติ วิธีนี้ให้ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นไม่ต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนดและให้ค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่นต่ำที่สุด เมื่อค่าพารามิเตอร์ θ เท่ากับ 1 ในทุกระดับของขนาดตัวอย่าง

วิธีการประมาณด้วยรากของสมการกำลังสอง วิธีนี้ให้ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นไม่ต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนดและให้ค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่นต่ำที่สุด เมื่อขนาดตัวอย่างเท่ากับ 100 และ 40 สำหรับกรณีที่ร้อยละของค่าผิดปกติเท่ากับ 5 และ 10 ตามลำดับ

วิธีการประมาณด้วยช่วงแบบเบส์ วิธีนี้ให้ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นไม่ต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนดและให้ค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่นต่ำที่สุดในสองกรณีคือ กรณีที่ 1 ขนาดตัวอย่างเท่ากับ 20, 40 และ 200 เมื่อร้อยละของค่าผิดปกติ เท่ากับ 5 และกรณีที่ 2 ขนาดตัวอย่างเท่ากับ 20, 100 และ 200 เมื่อร้อยละของค่าผิดปกติเท่ากับ 10 ในเกือบทุกระดับของค่าพารามิเตอร์

วราฤทธิ์ พานิชกิจโกศลกุล (2552) ได้เปรียบเทียบประสิทธิภาพของช่วงความเชื่อมั่นของพารามิเตอร์ สำหรับตัวแบบตัวแบบอัตตสหสัมพันธ์อันดับที่หนึ่ง (AR1) เมื่อข้อมูลมีค่าผิดปกติและค่าสูญหาย ช่วงความเชื่อมั่นที่พิจารณามี 3 วิธี คือ ช่วงความเชื่อมั่นโดยใช้วิธีกำลังสองน้อยที่สุดแบบค่าเฉลี่ยเวียนเกิด (RM) วิธีกำลังสองน้อยที่สุดแบบมัธยฐานเวียนเกิด (RMD) และวิธีกำลังสองน้อยที่สุดแบบมัธยฐานเวียนเกิดปรับปรุง (IRMD) ร้อยละของค่าผิดปกติ เท่ากับ 10 ขนาดของค่าผิดปกติ เท่ากับ $3\sigma_a$ และ $5\sigma_a$ ร้อยละของค่าสูญหาย เท่ากับ 10 กำหนดขนาดตัวอย่าง 4 ระดับ คือ 25, 50,100 และ 250 และคำนวณช่วงความเชื่อมั่น ที่ระดับความเชื่อมั่น 95% ในการวิจัยครั้งนี้ใช้วิธีการจำลองแบบมอนติคาร์โล และทำการทดลองซ้ำๆ กัน 10,000 ครั้ง ในแต่ละสถานการณ์ สำหรับเกณฑ์ที่ใช้ในการเปรียบเทียบประสิทธิภาพของช่วงความเชื่อมั่นพิจารณาจากค่าความน่าจะเป็นครอบคลุม (Coverage Probability) และค่าเฉลี่ยความกว้างช่วง (Expected Length) สรุปได้ดังนี้ ช่วงความเชื่อมั่นโดยใช้วิธี IRMD ให้ค่าความน่าจะเป็นครอบคลุมมากกว่าช่วงความเชื่อมั่นโดยใช้วิธี RM และ RMD ในเกือบทุกกรณีที่ศึกษา ยกเว้นกรณีที่ค่าพารามิเตอร์ ρ เท่ากับ 0.1 นอกจากนี้ ค่าความน่าจะเป็นครอบคลุมของช่วงความเชื่อมั่นทั้ง 3 วิธี มีค่าต่ำกว่า 0.95 ในเกือบทุกกรณีที่ศึกษา

งานวิจัยที่เกี่ยวข้องกับการประมาณอัตราส่วนประชากร

H. Toutenburg และ V.K. Srivastava (1998) ได้นำเสนอเกี่ยวกับตัวประมาณอย่างง่ายสำหรับอัตราส่วนของค่าเฉลี่ยประชากร 4 ตัว คือ $r_1 = \frac{\bar{y}}{\bar{x}}$, $r_2 = \frac{(n-q)\bar{y}}{(n-p-q)\bar{x} + p\bar{x}^*}$, $r_3 = \frac{(n-p-q)\bar{y} + q\bar{y}^{**}}{(n-p)\bar{x}}$ และ $r_4 = \left(\frac{n-q}{n-p}\right)\frac{(n-p-q)\bar{y} + q\bar{y}^{**}}{(n-p)\bar{x} + p\bar{x}^*}$ เมื่อข้อมูลบางค่าสูญหาย

โดยการประมาณค่า ของตัวอย่างขนาดใหญ่และได้เปรียบเทียบตัวประมาณค่าที่สร้างขึ้น ด้วยวิธีความสัมพันธ์เอนเอียงและความคลาดเคลื่อนกำลังสองเฉลี่ย ผลการศึกษาพบว่าตัวประมาณ r_2 และ r_3 มีประสิทธิภาพมากกว่า r_1 เมื่อสัมประสิทธิ์สหสัมพันธ์ของประชากรระหว่างตัวแปร X และตัวแปร Y มีทิศทางตรงกันข้ามกันหรือเป็นศูนย์ ถ้าสัมประสิทธิ์สหสัมพันธ์ของประชากรระหว่างตัวแปร X และตัวแปร Y มีทิศทางเดียวกัน r_2 จะมีประสิทธิภาพมากกว่า r_1 เมื่อ $\rho < \frac{\theta}{2}$ ขณะที่ r_3 มีประสิทธิภาพมากกว่า r_1 เมื่อ $\rho < \frac{1}{2\theta}$ ส่วน r_2 มีประสิทธิภาพมากกว่า r_3 เมื่อ $\rho < \Phi$ ถ้า $f_q < f_p$ และ $\rho > \Phi$ ถ้า $f_q > f_p$ สุดท้าย r_4 มีประสิทธิภาพมากกว่า r_2 เมื่อ $\rho\left(f_q - f_p\right) < \left(\frac{f_{p+q} - f_p}{2\theta}\right)$

Danute Krapavickaite และ Aleksandras Plikusas(2004) ได้ศึกษาการประมาณอัตราส่วนระหว่างผลรวมสองตัว เมื่อตัวอย่างความน่าจะเป็นมาจากประชากรที่จำกัด และได้เสนอตัวประมาณขึ้น 4 ตัว คือ (1) $\hat{R} = \frac{\hat{\tau}_y}{\hat{\tau}_z}$ (2) $\hat{R}^{(rat)} = \frac{\tau_{xy}\hat{\tau}_y\hat{\tau}_{xz}}{\tau_{xz}\hat{\tau}_z\hat{\tau}_{xy}}$ (3) $\hat{R}^{(reg)} = \frac{\hat{\tau}_y\left(\tau_{xy} - \hat{\tau}_{xy}\right)\hat{B}_y}{\hat{\tau}_z\left(\tau_{xz} - \hat{\tau}_{xz}\right)\hat{B}_z}$ และ (4) $\hat{R}^{(cal)} \approx R + \hat{R}_1^{(cal)} + \hat{R}_2^{(cal)}$ และได้เปรียบเทียบตัวประมาณทั้ง 4 ตัวนี้ด้วยค่าความคลาดเคลื่อนกำลังสองเฉลี่ยและค่าความแปรปรวนโดยประมาณ พบว่า $\hat{R}^{(cal)}$ มีค่าความคลาดเคลื่อนกำลังสองเฉลี่ยน้อยที่สุด