บทที่ 2

ทฤษฎีที่เกี่ยวข้องกับการวิจัย

ในการวิจัยครั้งนี้สิ่งที่สนใจศึกษาคือ การเปรียบเทียบวิธีการประมาณค่าสัดส่วนประชากรแบบช่วง ซึ่งประกอบด้วย วิธีการประมาณด้วยการแจกแจงแบบปกติเมื่อไม่ใช้ค่าปรับแก้ไขเพื่อความต่อเนื่อง วิธีการประมาณด้วยการแจกแจงแบบปกติเมื่อใช้ค่าปรับแก้ไขเพื่อความต่อเนื่อง และวิธีการประมาณด้วยการแจกแจงแบบเอฟ โดยวิธีการประมาณด้วยการแจกแจงแบบปกติทั้งกรณีที่ใช้และไม่ใช้ค่าปรับแก้ไขเพื่อความต่อเนื่อง จะพิจารณาเปรียบเทียบในลักษณะที่ผสมผสานกับค่าความคลาดเคลื่อนมาตรฐานอีก 2 ค่า รวมเป็นการศึกษาเปรียบเทียบทั้งหมด 5 วิธี ทั้งนี้การศึกษาเปรียบเทียบ จะพิจารณาจากค่าระดับความเชื่อมั่นของช่วงความเชื่อมั่น และค่าความยาวเฉลี่ยของช่วงความเชื่อมั่นที่คำนวณได้จากแต่ละวิธี ดังรายละเอียดในวัตถุประสงค์ของการวิจัยที่ได้กล่าวไปแล้วในบทที่ 1 ในบทนี้จะกล่าวถึงรายละเอียดของแต่ละวิธีการประมาณ ตลอดจนทฤษฎีอื่น ๆ ที่เกี่ยวข้องกับการวิจัย โดยมีรายละเอียดต่าง ๆ ดังนี้

2.1 การประมาณค่าพารามิเตอร์แบบช่วง

ให้ $X_1, X_2, \ldots, X_n$ เป็นตัวอย่างสุ่มจากการแจกแจงซึ่งมี θ เป็นพารามิเตอร์ที่ไม่ทราบค่า

สมมติสามารถหาตัวสถิติ $A(X_1, \ldots, X_n)$ และ $B(X_1, \ldots, X_n)$ ซึ่งสำหรับค่าจริง θ ใด ๆ

$$P(\ A(X_1, \ldots, X_n) < \theta < B(X_1, \ldots, X_n)\ ) = 1-\alpha$$

โดยที่ $1-\alpha$ เป็นความน่าจะเป็นคงที่ $(0 < \alpha < 1)$

จากนี้เมื่อทราบค่าของ $X_i = x_i$ $(i = 1, \ldots, n)$ และค่าของ $A(X_1, \ldots, X_n)$ และ $B(X_1, \ldots, X_n)$ สมมติให้เป็น a และ b ตามลำดับ ดังนั้นจะได้ช่วง (a, b)

และเรียกช่วง (a, b) ว่า " ช่วงความเชื่อมั่น 100(1-α) เปอร์เซนต์สำหรับ θ " [100(1-α)% confidence interval for θ] หรือกล่าวว่า ค่าจริงของ θ จะตกอยู่ในช่วง (a, b) ด้วยความเชื่อมั่น 100(1-α) เปอร์เซนต์ และเรียกค่า 1-α ว่า "สัมประสิทธิ์ความเชื่อมั่น " (confidence coefficient)

2.2 การแจกแจงของค่าสัดส่วนตัวอย่าง

ในการทดลองสุ่มใด ๆ ก็ตามที่มีผลลัพท์เป็นไปได้ 2 อย่าง คือ ความสำเร็จกับความไม่สำเร็จด้วยความน่าจะเป็น p และ q=(1-p) ตามลำดับ การทดลองนี้เรียกว่าการทดลองแบบแบร์นูลลี่ (Bernoulli experiment) ถ้าให้ X=1 เมื่อเกิดความสำเร็จ และ X=0 เมื่อเกิดความไม่สำเร็จ ตัวแปรสุ่ม X จะมีการแจกแจงความน่าจะเป็นแบบแบร์นูลลี่ (Bernoulli probability distribution) ด้วยพารามิเตอร์ p และมีฟังก์ชันความน่าจะเป็นในรูป

$$f(x;p) = p^{x}(1-p)^{1-x} \qquad ; x=0,1$$

ค่าความน่าจะเป็น p คือสัดส่วนประชากรที่เกิดความสำเร็จ ถ้าตัวอย่างสุ่มขนาด n ที่เลือกมาจากประชากรนี้ มีตัวอย่างสุ่มคือ $X_1, X_2, \ldots, X_n$ ฟังก์ชันที่ได้จากตัวอย่างสุ่มที่เราพิจารณา อาจจะเป็นสัดส่วนตัวอย่างที่เกิดความสำเร็จ คือ $\hat{P} = X/n = 1/n.\sum_{i=1}^{n} X_i$ เมื่อ $X = \sum_{i=1}^{n} X_i$ จะได้ว่า X คือจำนวนหน่วยที่เกิดความสำเร็จในตัวอย่างสุ่มขนาด n และเป็นตัวแปรสุ่มทวินามที่มีพารามิเตอร์เป็น n และ p โดยมีค่าเฉลี่ยเป็น np และความแปรปรวน npq แต่สำหรับตัวแปรสุ่ม $\hat{P}$ นี้ไม่ได้มีการแจกแจงในรูปแบบที่เรารู้จักกันดี อย่างไรก็ตาม ถ้าตัวอย่างสุ่มมีขนาดใหญ่พอ จะสามารถประยุกต์ทฤษฎีลิมิตสู่ส่วนกลางกับกรณีนี้ได้ โดยจะพิจารณาว่า $\hat{P} = 1/n.\sum_{i=1}^{n} X_i = \bar{X}$ เป็นค่าเฉลี่ยตัวอย่าง และได้ว่า

$$E(\hat{P}) = 1/n.\sum_{i=1}^{n} E(X_i) = 1/n.\sum_{i=1}^{n} p = p$$

$$V(\hat{P}) = V(\bar{X}) = V(X)/n^2 = pq/n$$

ดังนั้นถ้าตัวอย่างมีขนาดใหญ่พอ โดยทฤษฎีลิมิตสู่ส่วนกลาง $\hat{P}$ จะมีการแจกแจงโดยประมาณแบบปกติ โดยมีค่าเฉลี่ยของการแจกแจงเป็น p และค่าแปรปรวนเป็น pq/n เมื่อ q=1-p

2.3 ความสัมพันธ์ระหว่างการแจกแจงแบบเอฟกับการแจกแจงแบบเบตา [1]

ในการประมาณค่าสัดส่วนประชากรแบบช่วงด้วยการแจกแจงแบบเอฟนั้น จะใช้ค่าของตัวแปรสุ่มเอฟเป็นพื้นฐานในการสร้าง และการหาค่าของตัวแปรสุ่มเอฟ จะอาศัยความสัมพันธ์ระหว่างฟังก์ชันการแจกแจงแบบเอฟ กับฟังก์ชันการแจกแจงแบบเบตา ดังนี้

ถ้า F เป็นตัวแปรสุ่มแบบเอฟที่มีพารามิเตอร์ (m,n) จะหาค่าของตัวแปรสุ่มเอฟได้จาก

$$f = \frac{nx}{m - mx}$$

ซึ่ง x คือค่าที่สอดคล้องกับสมการ

$$P(F \leqslant f) = I_x(m/2,n/2) = \frac{1}{B(m/2,n/2)} \int_0^x t^{m/2-1} (1-t)^{n/2-1}\, dt$$

---

[1] Johnson N.L. and Kotz S., Continuous Univariate Distribution-2, New York : John Wiley & Sons, 1970, pp. 78.

2.4 ผลงานที่เกี่ยวข้องกับการวิจัย

รายละเอียดที่จะนำเสนอต่อไป เป็นผลงานของนักสถิติในรุ่นก่อน ที่ได้ถ่ายทอดมาสู่ปัจจุบัน โดยจะเป็นผลงานในส่วนที่เกี่ยวกับ วิธีการประมาณช่วงความเชื่อมั่น 2 วิธีใหญ่ ๆ คือการประมาณโดยใช้การแจกแจงแบบปกติ ทั้งกรณีที่ใช้และไม่ใช้ค่าปรับแก้ไขเพื่อความต่อเนื่อง และการประมาณโดยใช้การแจกแจงแบบเอฟ ตลอดจนรายละเอียดเกี่ยวกับค่าความคลาดเคลื่อนมาตรฐาน ที่ใช้ในสูตรของการประมาณค่าสัดส่วนประชากรแบบช่วงด้วยการแจกแจงแบบปกติ

2.4.1 การประมาณโดยใช้การแจกแจงแบบปกติ

กฎเลขจำนวนมากของแบร์นูลลี่

สำหรับการทดลองแบบแบร์นูลลี่ ที่มีความน่าจะเป็นของการเกิดความสำเร็จเท่ากับ p กำหนดให้

S_n แทนจำนวนครั้งของการเกิดความสำเร็จทั้งหมด จากการทดลองแบบแบร์นูลลี่ที่เป็นอิสระกัน n ครั้ง ดังนั้น $S_n \sim b(n,p)$

f_n เป็นค่าความถี่สัมพัทธ์ของจำนวนครั้งที่เกิดความสำเร็จ ในการทดลองแบบแบร์นูลลี่ n ครั้ง และหาได้จากความสัมพันธ์

$$f_n = S_n/n$$

สำหรับจำนวนเต็มบวก $\epsilon > 0$ ใด ๆ ที่กำหนดให้ จะได้ว่า

$$\lim_{n\to\infty} P(\ |f_n - p| < \epsilon\ ) = 1$$

หรือ $$\lim_{n\to\infty} P(\ |f_n - p| \geq \epsilon\ ) = 0$$

กฎเลขจำนวนมากของแบร์นูลลี่ กล่าวว่า ในการประมาณค่า p ซึ่งเป็นค่าความน่าจะเป็นของการเกิดความสำเร็จในการทดลองแบบแบร์นูลลี่ จะต้องทำการทดลองแล้วสังเกตจำนวนครั้งที่เกิดความสำเร็จในการทดลอง ค่าประมาณที่ได้จะมีความถูกต้องใกล้กับค่าที่แท้จริงของ p เมื่อจำนวนครั้งของการทดลองเข้าใกล้อนันต์

ในปี ค.ศ. 1733 อะบราฮัม เดอร์มัวฟ์ (Abraham De Moivre: ค.ศ.1667-1754) ได้เสนอผลงานซึ่งเป็นการประมาณฟังก์ชันการแจกแจงแบบทวินามที่มีพารามิเตอร์ (n,p=1/2) ด้วยฟังก์ชันการแจกแจงของค่าผิดพลาดจากการวัด (Distribution of Error) ซึ่งต่อมาเป็นที่รู้จักกันในชื่อการแจกแจงแบบปกติ (Normal distribution)

ต่อมาในปี ค.ศ. 1812 ปิแยร์ ซิมง เดอ ลาปลาซ (Pierre Simon de Laplace: ค.ศ. 1749-1827) ได้เสนอผลงานซึ่งเป็นการขยายความคิดของเดอร์มัวฟ์ในการประมาณฟังก์ชันการแจกแจงแบบทวินาม ที่มีพารามิเตอร์ (n,p) เมื่อ p ไม่เท่ากับ 1/2

ผลงานของท่านทั้งสองต่อมาเป็นที่รู้จักในชื่อ " Demoivre - Laplace Limit Theorem " ซึ่งมีใจความโดยย่อดังต่อไปนี้

<u>ทฤษฎีลิมิตของเดอร์มัวฟ์-ลาปลาซ</u>

สำหรับการทดลองแบบแบร์นูลลี่ ที่มีความน่าจะเป็นของการเกิดความสำเร็จเท่ากับ p กำหนดให้ S_n เป็นจำนวนครั้งของผลสำเร็จในการทดลองแบบแบร์นูลลี่ n ครั้งที่เป็นอิสระกัน จะได้ว่าเมื่อ n เข้าใกล้อนันต์ ตัวแปรสุ่ม S_n จะมีการแจกแจงลู่เข้าสู่การแจกแจงแบบปกติ ที่มีค่าเฉลี่ยเท่ากับ np และค่าความแปรปรวนเท่ากับ np(1-p)

จากทฤษฎีลิมิตของ เดอร์มัวฟ์-ลาปลาซ ได้นำไปสู่สูตรการประมาณการแจกแจงความน่าจะเป็นแบบทวินามโดยใช้การแจกแจงแบบปกติ

$$\sum_{k=a}^{b}\binom{n}{k}p^k q^{n-k} \doteq \frac{1}{\sqrt{2\pi}}\int_{\frac{a-np-\frac{1}{2}}{\sqrt{npq}}}^{\frac{b-np+\frac{1}{2}}{\sqrt{npq}}} e^{-\frac{1}{2}y^2}\,dy$$

$$= \Phi\left(\frac{b-np+\frac{1}{2}}{\sqrt{npq}}\right) - \Phi\left(\frac{a-np-\frac{1}{2}}{\sqrt{npq}}\right)$$

จากสูตรการประมาณการแจกแจงความน่าจะเป็นแบบทวินามโดยใช้การแจกแจงแบบปกติ เราจะหาสูตรการประมาณค่าสัดส่วนประชากรแบบช่วง ของการแจกแจงแบบทวินามได้ดังนี้

เมื่อ $\hat{P}$ คือตัวประมาณแบบจุดของ p จาก

$$Z = \frac{\hat{P} - p}{\sqrt{p(1-p)/n}} \sim N(0,1)$$

และ

$$P(Z\alpha/2 < Z < Z1-\alpha/2) = 1-\alpha$$

$$P\left(-Z1-\alpha/2 < \frac{\hat{P} - p}{\sqrt{p(1-p)/n}} < Z1-\alpha/2\right) = 1-\alpha$$

$$P\left(\hat{P} - Z1-\alpha/2\sqrt{p(1-p)/n} < p < \hat{P} + Z1-\alpha/2\sqrt{p(1-p)/n}\right) = 1-\alpha$$

ได้ว่าช่วงความเชื่อมั่น $(1-\alpha)100\%$ สำหรับ p คือ

$$\left(\hat{p} - Z1-\alpha/2\sqrt{p(1-p)/n}\ ,\ \hat{p} + Z1-\alpha/2\sqrt{p(1-p)/n}\right)$$

แต่เนื่องจากไม่ทราบค่าสัดส่วนประชากร p ฉะนั้นจะประมาณค่า p และวิธีหนึ่งที่

ใช้กันทั่วไป คือ ประมาณค่า p ด้วยค่าสัดส่วนจากตัวอย่าง $\hat{p}$ ดังนั้นช่วงความเชื่อมั่น $(1-\alpha)100\%$ สำหรับ p จะเป็นดังนี้

$$(\hat{p} - Z1\text{-}\alpha/2 \sqrt{\hat{p}(1-\hat{p})/n} \ , \ \hat{p} + Z1\text{-}\alpha/2 \sqrt{\hat{p}(1-\hat{p})/n})$$

เมื่อ $\hat{p} = x/n$

x คือ ค่าของตัวแปรสุ่มทวินามที่สังเกตได้

n คือ ค่าขนาดตัวอย่าง

$Z1\text{-}\alpha/2$ คือ ค่าของตัวแปรสุ่มปกติมาตรฐาน ที่ให้ค่าของฟังก์ชันความน่าจะเป็นสะสมเท่ากับ $1-\alpha/2$

2.4.2 ค่าความคลาดเคลื่อนมาตรฐานของสัดส่วนตัวอย่าง

ในการประมาณช่วงความเชื่อมั่นสำหรับค่าสัดส่วนประชากร ด้วยการแจกแจงแบบปกตินั้น จากรูปแบบของช่วงที่ได้ คือ

$$(\hat{p} - Z1\text{-}\alpha/2 \sqrt{\hat{p}(1-\hat{p})/n} \ , \ \hat{p} + Z1\text{-}\alpha/2 \sqrt{\hat{p}(1-\hat{p})/n})$$

ค่าความแปรปรวน $s^2 = \hat{p}(1-\hat{p})/n$ เป็นค่าประมาณที่เอนเอียงทางด้านลบ (underestimate) ของค่าความแปรปรวนของค่าสัดส่วนประชากร ฉะนั้นในกรณีของตัวอย่างขนาดเล็ก การประมาณค่าสัดส่วนประชากรแบบช่วงโดยใช้ค่าความคลาดเคลื่อนมาตรฐาน $s = \sqrt{\hat{p}(1-\hat{p})/n}$ จะทำให้ได้ช่วงของการประมาณที่แคบเกินไป ดังนั้นจึงควรปรับค่าความคลาดเคลื่อนมาตรฐาน ซึ่งได้มีผู้เสนอให้ใช้ค่าความคลาดเคลื่อนมาตรฐานเป็น $s = \sqrt{\hat{p}(1-\hat{p})/(n-1)}$ เนื่องจาก $s^2 = \hat{p}(1-\hat{p})/(n-1)$ เป็นค่าประมาณที่ไม่เอนเอียงของความแปรปรวนของค่าสัดส่วนประชากร ดังนั้น จะได้ช่วงความเชื่อมั่นเป็น

$$(\hat{p} - Z1\text{-}\alpha/2 \sqrt{\hat{p}(1-\hat{p})/(n-1)} \ , \ \hat{p} + Z1\text{-}\alpha/2 \sqrt{\hat{p}(1-\hat{p})/(n-1)})$$

2.4.3 ค่าปรับแก้ไขเพื่อความต่อเนื่อง

เมื่อจำนวนครั้งของการทดลองมีขนาดใหญ่ จากที่ได้กล่าวมาแล้วว่า สามารถประมาณการแจกแจงความน่าจะเป็นแบบทวินามได้ด้วยการแจกแจงแบบปกติ แต่เนื่องจากการแจกแจงความน่าจะเป็นแบบทวินามเป็นการแจกแจงของตัวแปรสุ่มชนิดไม่ต่อเนื่อง (Discrete Random Variable) ในขณะที่ การแจกแจงแบบปกติเป็นการแจกแจงชนิดที่ต่อเนื่อง (Continuous Random Variable) ดังนั้น ในการประมาณค่าสัดส่วนประชากรแบบช่วงด้วยการแจกแจงแบบปกติ จึงควรปรับช่วงของความน่าจะเป็น เพื่อให้ได้ความน่าจะเป็นใกล้ค่าจริงหรือมีความคลาดเคลื่อนน้อย ค่าที่ใช้ปรับแก้ไขเรียกว่า ค่าปรับแก้ไขเพื่อความต่อเนื่อง (continuity correction)

ค่าปรับแก้ไขเพื่อความต่อเนื่องที่มีผู้เสนอไว้และปรากฏใช้ในหนังสือหลายเล่ม คือ ค่าปรับแก้ไขเพื่อความต่อเนื่องของเยจส์ (Yates's correction for continuity) ซึ่งมีค่าเท่ากับ 1/(2n) ดังนั้นจะได้ช่วงความเชื่อมั่นของค่าสัดส่วนประชากร ที่ประมาณด้วยการแจกแจงแบบปกติ เมื่อใช้ค่าปรับแก้ไขเพื่อความต่อเนื่องของเยจส์ ที่ระดับความเชื่อมั่น $(1-\alpha)100\%$ คือ

$$\left(\hat{p} - (2n)^{-1} - Z_{1-\alpha/2} \sqrt{\hat{p}(1-\hat{p})/n}\ ,\ \hat{p} + (2n)^{-1} + Z_{1-\alpha/2} \sqrt{\hat{p}(1-\hat{p})/n} \right)$$

2.4.4 การประมาณโดยใช้การแจกแจงแบบเอฟ

ในปี ค.ศ. 1924 คาร์ล เพียร์สัน [2] ได้พิสูจน์ความสัมพันธ์ระหว่าง ผลบวก p เทอมแรกของการกระจายทวินาม $(a + b)^n$ กับฟังก์ชันเบตาที่ไม่สมบูรณ์ (Incomplete Beta function) ซึ่งสรุปได้ดังนี้

---

[2] Karl Pearson, Note on the Relationship of the Incomplete B-function to the Sum of the first p terms of the Binomial $(a + b)^n$ (Biometika, Vol. 16, 1924), pp. 202-203.

สำหรับการกระจายทวินาม $(a + b)^n$ สามารถกระจายออกเป็น

$$(a + b)^n = C_{0,n}a^0b^n + C_{1,n-1}a^1b^{n-1} + \dots + C_{n,0}a^nb^0$$

ซึ่ง $C_{i,j} = n!/(i!j!)$

ดังนั้นผลบวก p เทอมแรกของการกระจาย คือ

$$C_{0,n}a^0b^n + C_{1,n-1}a^1b^{n-1} + \dots + C_{p-1,n-p+1}\,a^{p-1}b^{n-p+1}$$

เพียร์สันได้แสดงว่า ผลบวก p เทอมแรกดังกล่าว มีค่าเท่ากับ

$$(a + b)^n\left[1 - \frac{B_x(p, q)}{B(p, q)}\right]$$

เมื่อ $x = b/(a + b)$

$q = n - p + 1$

$B(p, q)$ คือ ฟังก์ชันเบตา (Bata function) ซึ่งกำหนดโดย

$$B(p, q) = \int_0^1 t^{p-1}(1 - t)^{q-1}\,dt$$

$B_x(p, q)$ คือฟังก์ชันเบตาที่ไม่สมบูรณ์ (Incomplete Beta function) กำหนดโดย

$$B_x(p, q) = \int_0^x t^{p-1}(1 - t)^{q-1}\,dt \quad ; \quad 0 \leqslant x \leqslant 1$$

ต่อมาในปี ค.ศ. ทอมสัน 1942 [3] ได้สร้างตารางการแจกแจงเบตาขึ้น โดยในตอนนั้นเรียกว่า อัตราส่วนของฟังก์ชันเบตาที่ไม่สมบูรณ์ (Incomplete Beta function ratio) แทนด้วย

$$I_x(p, q) = \frac{B_x(p, q)}{B(p, q)} \quad ; \quad 0 \leqslant x \leqslant 1$$

$$= \frac{1}{B(p, q)} \int_0^x t^{p-1} (1-t)^{q-1} dt$$

ซึ่งจะเห็นว่า เทอมทางขวามือของสมการดังกล่าว คือ ฟังก์ชันการแจกแจงแบบเบตาที่รู้จักกันในปัจจุบันนั่นเอง

ในปี ค.ศ. 1952 ฮาล [4] ได้พิสูจน์ความสัมพันธ์ระหว่าง ฟังก์ชันการแจกแจงแบบทวินาม กับ ฟังก์ชันการแจกแจงแบบเบตา * ดังนี้

สำหรับตัวแปรสุ่ม X ซึ่งมีการแจกแจงแบบทวินาม ที่มีพารามิเตอร์ (n,p) จะได้ว่า

$$P(X \geqslant x) = B(x, n - x + 1)^{-1} \int_0^p t^{x-1} (1-t)^{n-x} dt$$

---

[3] Catherine M. Thompson, Tables of Percentage Points of the Incomplete BETA-function Biometrika, Vol. 32, 1941-1942, pp.151-181.

[4] A. Hald, Statistical Theory with Engineering Applications New York: John Wiley and Sons, 1952, pp. 674.

* ขณะนั้นยังคงเรียกว่า Incomplete Beta function ratio

$$= I_p(x,\ n - x + 1)$$

และ $$P(X \leq x) = I_{1-p}(n - x,\ x + 1)$$

โดยที่

$$I_{1-p}(n - x,\ x + 1) = 1 - I_p(x + 1,\ n - x)$$

และ ฮาล ยังได้เสนอความสัมพันธ์ระหว่าง ฟังก์ชันการแจกแจงแบบทวินาม กับ ฟังก์ชันการแจกแจงแบบเอฟ ดังนี้

สำหรับตัวแปรสุ่ม X ซึ่งมีการแจกแจงแบบทวินาม ที่มีพารามิเตอร์ (n,p) จะได้ว่า

$$P(X \leq x) = 1 - P\left(F2 < \frac{n - x}{x + 1}\left(\frac{p}{1 - p}\right)\right)$$

เมื่อ F2 เป็นตัวแปรสุ่มเอฟ ที่มีองศาอิสระ (2[x + 1], 2[n - x]) และจะได้ว่า

$$P(X \geq x) = P\left(F1 < \frac{n - x + 1}{x}\left(\frac{p}{1 - p}\right)\right)$$

เมื่อ F1 เป็นตัวแปรสุ่มเอฟ ที่มีองศาอิสระ (2[n - x + 1], 2x)

จากความสัมพันธ์ดังกล่าว จะสามารถสร้างช่วงความเชื่อมั่น (1-α)100% สำหรับค่าสัดส่วนประชากร p ของการแจกแจงทวินาม เป็นช่วง (PL, PU) ซึ่งสอดคล้อง

$$P(PL < p < PU) = 1 - \alpha$$

ฮาล ได้แสดงว่า สามารถหาค่า PL และ PU ได้จากสมการ

$$\sum_{k=x}^{n} C_{n,k} (PL)^{k} (1 - PL)^{n-k} = \alpha/2$$

$$\sum_{k=0}^{x} C_{n,k} (PU)^{k} (1 - PU)^{n-k} = \alpha/2$$

ซึ่งได้คำตอบเป็น

$$PL = \frac{x}{x + (n - x + 1)F1}$$

$$PU = \frac{(x + 1)F2}{(x + 1)F2 + (n - x)}$$

ซึ่ง

x คือ จำนวนครั้งของผลสำเร็จในตัวอย่าง

n คือ ขนาดตัวอย่าง

F1 คือ ค่าตัวแปรสุ่มเอฟ ที่ให้ค่าของฟังก์ชันความน่าจะเป็นสะสมเท่ากับ $1-\alpha/2$ และมีค่าองศาอิสระ [2(n-x+1),2x]

F2 คือ ค่าตัวแปรสุ่มเอฟ ที่ให้ค่าของฟังก์ชันความน่าจะเป็นสะสมเท่ากับ $1-\alpha/2$ และมีค่าองศาอิสระ [2(x+1),2(n-x)]