บทที่ 2

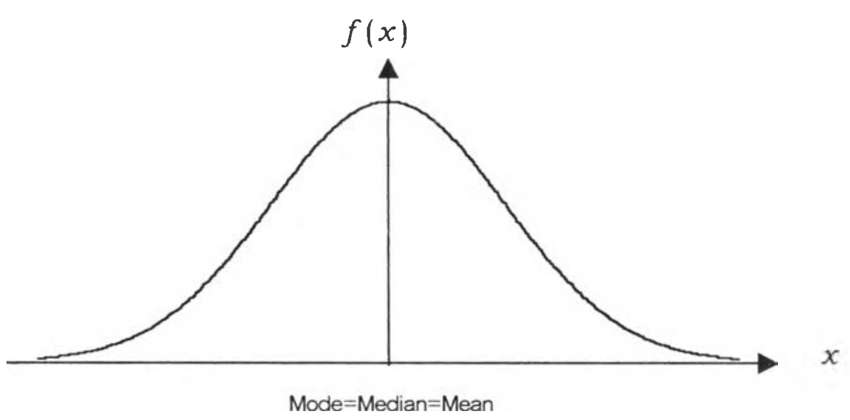
ทฤษฎีและสถิติที่เกี่ยวข้องกับการวิจัย

การประมาณค่าแบบช่วงที่เหมาะสมนั้น ช่วงความเชื่อมั่นที่ประมาณได้จะต้องครอบคลุมค่าพารามิเตอร์ที่สนใจด้วยระดับความเชื่อมั่นที่กำหนดและช่วงความเชื่อมั่นที่ประมาณได้นั้นควรเป็นช่วงที่แคบ ดังนั้นในงานวิจัยนี้จึงมีจุดมุ่งหมายเพื่อเปรียบเทียบวิธีการประมาณค่าแบบช่วงสำหรับค่าเฉลี่ยของประชากรที่มีการแจกแจงแบบเบ้ขวา โดยจะพิจารณาการแจกแจงของประชากร 4 การแจกแจง คือ การแจกแจงไคกำลังสอง การแจกแจงลอกนอร์มอล การแจกแจงแกมมา และการแจกแจงไวบูลล์ ภายใต้ระดับความเบ้ต่างๆ ส่วนเกณฑ์ในการพิจารณานั้นจะพิจารณาค่าสัมประสิทธิ์ความเชื่อมั่น ค่าเฉลี่ยของขีดจำกัดความเชื่อมั่นล่าง (mean of lower limit) ค่าเฉลี่ยของขีดจำกัดความเชื่อมั่นบน (mean of upper limit) และค่าเฉลี่ยความยาวของช่วงความเชื่อมั่น (mean of confidence interval length) สำหรับในบทนี้จะกล่าวถึงทฤษฎีต่างๆที่เกี่ยวข้องกับการวิจัย เกณฑ์ในการตัดสินใจและผลงานวิจัยที่เกี่ยวข้องซึ่งมีรายละเอียดต่างๆดังนี้

2.1 ความเบ้ (Skewness) และความโด่ง (Kurtosis)

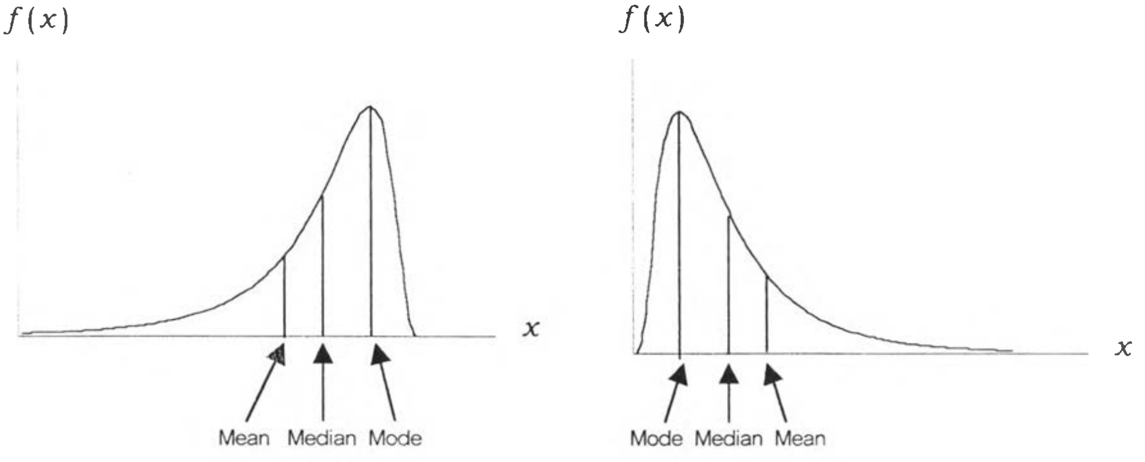
2.1.1 ความเบ้ (Skewness)

ประชากรที่มีการแจกแจงแบบสมมาตร.นั้น เส้นโค้งของการแจกแจงจะมีลักษณะเส้นทางด้านขวาของค่าเฉลี่ยและทางด้านซ้ายของค่าเฉลี่ยเหมือนกันทุกประการ ค่าเฉลี่ย (mean) ค่ามัธยฐาน (median) และค่าฐานนิยม (mode) จะมีค่าเท่ากันหรือทับกันสนิท ตัวอย่างเช่นเส้นโค้งของข้อมูลที่มีการแจกแจงแบบปกติดังรูปที่ 2.1



รูปที่ 2.1 แสดงเส้นโค้งของข้อมูลที่มีการแจกแจงแบบปกติ (รูป ก.)

ส่วนประชากรที่มีการแจกแจงแบบไม่สมมาตร เส้นโค้งที่ได้จากการแจกแจงจะมีลักษณะเบ้ไปข้างใดข้างหนึ่ง ค่าเฉลี่ย ค่ามัธยฐาน และค่าฐานนิยม จะมีค่าต่างกัน ดังรูปที่ 2.2



รูป ข. **รูป ค.**

รูปที่ 2.2 แสดงเส้นโค้งของข้อมูลที่มีการแจกแจงแบบไม่สมมาตร
(รูป ข. และ ค. แสดงการเบ้ซ้ายและเบ้ขวาของข้อมูล ตามลำดับ)

จะเห็นได้ว่าในรูป ข. ประชากรมีการแจกแจงเบ้ไปทางซ้ายเพราะพื้นที่ใต้เส้นโค้งทางด้านซ้ายของค่าฐานนิยมมีมากกว่าพื้นที่ใต้เส้นโค้งทางด้านขวาของฐานนิยม ในกรณีนี้ค่าเฉลี่ยจะมีค่าน้อยกว่าค่ามัธยฐานและค่ามัธยฐานจะมีค่าน้อยกว่าค่าฐานนิยม ส่วนในรูป ค.ประชากรมีการแจกแจงเบ้ไปทางขวาเพราะพื้นที่ใต้เส้นโค้งทางด้านขวาของฐานนิยมมีมากกว่าพื้นที่ใต้เส้นโค้งทางด้านซ้ายของฐานนิยม ในกรณีนี้ค่าเฉลี่ยจะมีค่ามากกว่าค่ามัธยฐานและค่ามัธยฐานจะมีค่ามากกว่าค่าฐานนิยม

การวัดความเบ้ (measure of skewness) **หรือการหาค่าสัมประสิทธิ์ความเบ้ จะใช้**การวัดความเบ้โดยวิธีโมเมนต์ (moment) **ซึ่งมีสูตรดังนี้**

$$\alpha_3 = \frac{\mu_3}{\sigma^3} = \frac{E\left[(X-\mu)^3\right]}{[Var(X)]^{3/2}}$$

เมื่อ μ_3 คือ โมเมนต์ศูนย์กลางอันดับที่ 3 $= E\left[(X-\mu)^3\right]$

และ σ คือ ส่วนเบี่ยงเบนมาตรฐานของประชากร $= \sqrt{Var(X)}$

การประมาณค่าสัมประสิทธิ์ความเบ้ของประชากรสามารถประมาณได้จากสัมประสิทธิ์ความเบ้ของตัวอย่าง ซึ่งมีสูตรดังนี้

$$\hat{\alpha}_3 = \frac{m_3}{(m_2)^{3/2}}$$

เมื่อ $m_3 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^3}{n}$

และ $m_2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n}$

การวัดความเบ้ด้วยโมเมนต์ศูนย์กลางอันดับ 3 จะให้ค่าต่างๆกันดังนี้

1. ถ้าการแจกแจงสมมาตร ค่าสัมประสิทธิ์ความเบ้มีค่าเป็นศูนย์
2. ถ้าการแจกแจงเบ้ขวา ค่าสัมประสิทธิ์ความเบ้มีค่าเป็นบวก
3. ถ้าการแจกแจงเบ้ซ้าย ค่าสัมประสิทธิ์ความเบ้มีค่าเป็นลบ

2.1.2 ความโด่ง (Kurtosis)

ในการพิจารณาลักษณะของการแจกแจง นอกจากจะพิจารณาความเบ้แล้ว ความโด่งเป็นสิ่งที่ใช้อธิบายการแจกแจงที่สำคัญอีกอย่างหนึ่ง ความโด่งของการแจกแจงของประชากรมี 3 ลักษณะ ดังนี้

ก) เส้นโค้งที่มีความโด่งเป็นปกติ เรียกว่า เส้นโค้งชนิด Mesokurtic

ข) เส้นโค้งที่แบนราบกว่าปกติ เรียกว่า เส้นโค้งชนิด Platykurtic

ค) เส้นโค้งที่โด่งมากกว่าปกติ เรียกว่า เส้นโค้งชนิด Leptokurtic

การวัดความโด่ง (measure of kurtosis) หรือการหาค่าสัมประสิทธิ์ความโด่ง จะหาได้จากค่าโมเมนต์ศูนย์กลางอันดับที่ 4 หารด้วยส่วนเบี่ยงเบนมาตรฐานยกกำลังสี่ ซึ่งมีสูตรดังนี้

$$\alpha_4 = \frac{\mu_4}{\sigma^4} = \frac{E\left[(X-\mu)^4\right]}{[Var(X)]^2}$$

เมื่อ μ_4 คือ โมเมนต์ศูนย์กลางอันดับที่ 4 $= E\left[(X-\mu)^4\right]$

และ σ คือ ส่วนเบี่ยงเบนมาตรฐานของประชากร $= \sqrt{Var(X)}$

การประมาณค่าสัมประสิทธิ์ความโด่งของประชากรสามารถประมาณได้จากสัมประสิทธิ์ความโด่งของตัวอย่าง ซึ่งมีสูตรดังนี้

$$\hat{\alpha}_4 = \frac{m_4}{(m_2)^2}$$

เมื่อ $m_4 = \dfrac{\sum_{i=1}^{n}(X_i - \overline{X})^4}{n}$

และ $m_2 = \dfrac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n}$

การวัดความโด่งด้วยโมเมนต์ศูนย์กลางอันดับ 4 จะให้ค่าต่างๆกันดังนี้

1. ถ้า $\alpha_4 = 3$ แสดงว่าเส้นโค้งที่มีความโด่งเป็นปกติ (Mesokurtic)
2. ถ้า $\alpha_4 < 3$ แสดงว่าเส้นโค้งที่แบนราบกว่าปกติ (Platykurtic)
3. ถ้า $\alpha_4 > 3$ แสดงว่าเส้นโค้งที่โด่งมากกว่าปกติ (Leptokurtic)

2.2 การแจกแจงต่าง ๆที่ใช้ในการวิจัย

2.2.1 <u>การแจกแจงไคกำลังสอง (Chi-Square Distribution)</u>

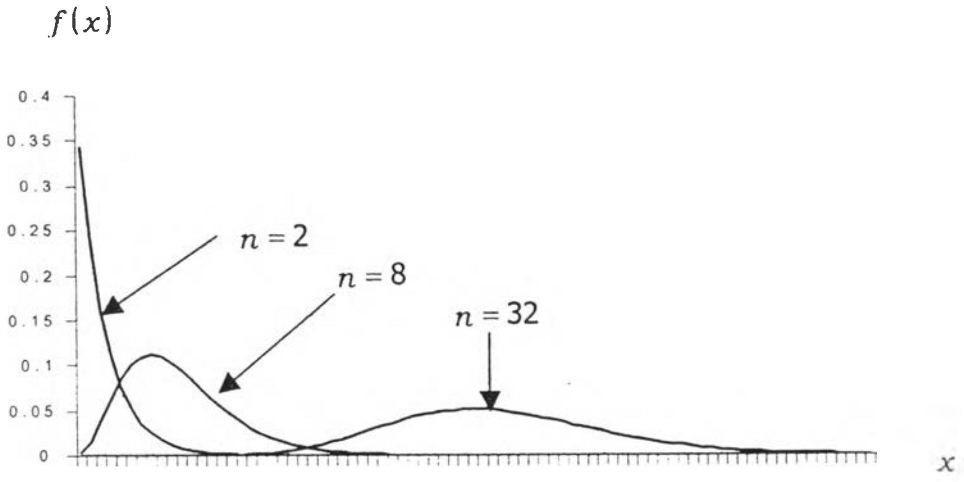
ในปี ค.ศ. 1876 เอฟ.อาร์.ฮิลเมอร์ท (F.R. Helmert) นักฟิสิกส์ชาวเยอรมันได้ค้นพบการแจกแจงไคกำลังสองเป็นครั้งแรก ต่อมาในปี ค.ศ. 1900 คาร์ล เพียร์สัน (Karl Pearson) นักสถิติชาวอังกฤษได้ทำการพัฒนาและนำมาใช้ประโยชน์อย่างแพร่หลาย

ให้ X เป็นตัวแปรสุ่มต่อเนื่องที่มีการแจกแจงไคกำลังสอง ด้วยระดับขั้นความเป็นเสรี (ร.ส.) = n (degree of freedom) ดังนั้นฟังก์ชันความหนาแน่นของ X จะอยู่ในรูปของ

$$f(x) = \frac{2^{-n/2}}{\Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-x/2} \qquad , x > 0$$

เมื่อ สัมประสิทธิ์ความเบ้เท่ากับ $\sqrt{\dfrac{8}{n}}$

และ สัมประสิทธิ์ความโด่งเท่ากับ $3 + \dfrac{12}{n}$



รูปที่ 2.3 **แสดงการแจกแจงไคกำลังสอง เมื่อ ร.ส. เท่ากับ** 2 , 8 **และ** 32

<u>คุณสมบัติของการแจกแจงไคกำลังสองมีดังนี้</u>

1. ค่าเฉลี่ยของการแจกแจงไคกำลังสองมีค่า = n (ร.ส.)
2. ค่ามัธยฐานของการแจกแจงไคกำลังสองมีค่าเท่ากับ $n-2 \quad , n \geq 2$
3. ค่าความแปรปรวนของการแจกแจงไคกำลังสองมีค่าเท่ากับ 2n
4. สัมประสิทธิ์ความเบ้มีค่าเป็นบวก ทำให้มีลักษณะการกระจายแบบเบ้ขวา
5. ค่า χ^2 มีค่ามากกว่าหรือเท่ากับศูนย์ เนื่องจากค่า $\chi_n^2 = Z_1^2 + Z_2^2 + \cdots + Z_n^2$ นั่นเอง เมื่อ Z_i มีการแจกแจง N(0,1) เหมือนกันและเป็นอิสระซึ่งกันและกัน
6. ถ้า n มีขนาดใหญ่แล้ว การแจกแจงไคกำลังสองจะมีลักษณะเข้าสู่การแจกแจงแบบปกติที่มีค่าเฉลี่ยเป็น n และความแปรปรวนเป็น 2n
7. ถ้า $\chi_{n_1}^2$ และ $\chi_{n_2}^2$ เป็นอิสระต่อกัน จะได้ว่า $\chi_{n_1}^2 + \chi_{n_2}^2$ มีการแจกแจง $\chi_{n_1+n_2}^2$ เมื่อ n_1, n_2 เป็น ร.ส. ของการแจกแจงไคกำลังสองชุดที่ 1 และชุดที่ 2 ตามลำดับ

2.2.2 <u>การแจกแจงลอกนอร์มอล (Log-normal Distribution)</u>

การแจกแจงลอกนอร์มอลเป็นการแจกแจงของตัวแปรสุ่มซึ่งมาจากการแจกแจงแบบปกติ การแจกแจงลอกนอร์มอลสามารถนำไปประยุกต์ใช้ได้อย่างกว้างขวาง เช่น วิทยาศาสตร์ สังคมศาสตร์ และวิศวกรรมศาสตร์

ให้ X เป็นตัวแปรสุ่มต่อเนื่องที่มีการแจกแจงลอกนอร์มอล ด้วยพารามิเตอร์ μ และ σ^2 ดังนั้นฟังก์ชันความหนาแน่นของ X จะอยู่ในรูปของ

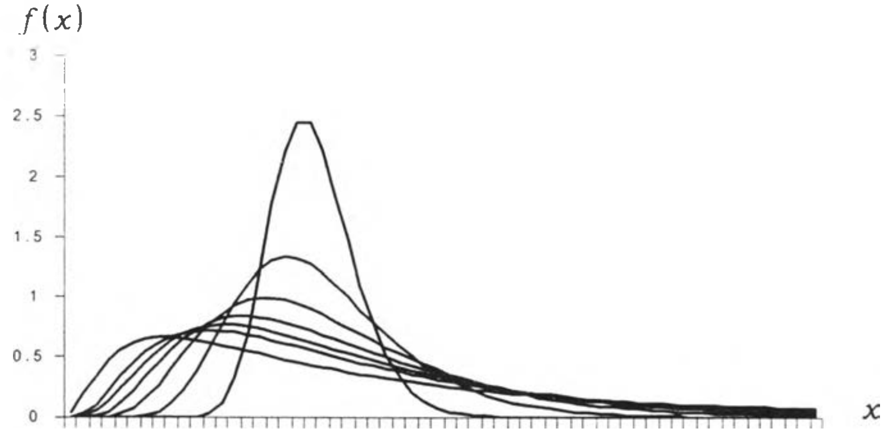
$$f(x)=\frac{1}{x\sigma\sqrt{2\pi}}\exp\left[-\frac{\{\ln(x)-\mu\}^2}{2\sigma^2}\right] \qquad ,x>0 \quad ,\sigma^2>0$$

เมื่อ μ และ σ^2 เป็นค่าเฉลี่ยและความแปรปรวนของ Y โดยที่ Y = ln X

Y มีการแจกแจงแบบปกติ

และ สัมประสิทธิ์ความเบ้เท่ากับ $(w+2)\sqrt{w-1}$

ส่วนสัมประสิทธิ์ความโด่งเท่ากับ $w^4+2w^3+3w^2-3$ โดยที่ $w=\exp(\sigma^2)$



รูปที่ 2.4 แสดงการแจกแจงลอกนอร์มอล เมื่อ $\mu=0$ และ
$\sigma=0.1641,0.3143,0.4435,0.5514,0.6409,0.7156,0.9202$
โดยค่าน้อยที่สุดมีความโด่งสูงสุดและลดหลั่นกัน ตามลำดับ

<u>คุณสมบัติของการแจกแจงลอกนอร์มอลมีดังนี้</u>

1. ค่าเฉลี่ย $=\exp\left\{\mu+\frac{\sigma^2}{2}\right\}$ และ

   ความแปรปรวน $=\exp(2\mu+\sigma^2)\{\exp(\sigma^2)-1\}$
2. เส้นโค้งมีลักษณะการกระจายแบบเบ้ขวา
3. ค่ามัธยฐาน $=\exp(\mu)$ และค่าฐานนิยม $=\exp(\mu-\sigma^2)$
4. ถ้า X มีการแจกแจงแบบลอกนอร์มอล ด้วยพารามิเตอร์ μ และ σ^2 กำหนดให้ a, b และ d เป็นค่าคงที่ ซึ่ง $b=\exp(d)$ แล้ว $W=bX^a$ จะมีการแจกแจงแบบลอกนอร์มอล ด้วยพารามิเตอร์ $(d+a\mu)$ และ $(a\sigma)^2$
5. ถ้า X_1 และ X_2 มีการแจกแจงแบบลอกนอร์มอลที่เป็นอิสระซึ่งกันและกัน ด้วยพารามิเตอร์ $(\mu_{Y_1},\sigma^2_{Y_1})$ และ $(\mu_{Y_2},\sigma^2_{Y_2})$ ตามลำดับ แล้ว $W=X_1X_2$ จะมีการแจกแจงแบบลอกนอร์มอล ด้วยพารามิเตอร์

   $\left(\mu_{Y_1}+\mu_{Y_2},\sigma^2_{Y_1}+\sigma^2_{Y_2}\right)$

2.2.3 <u>การแจกแจงแกมมา (Gamma Distribution)</u>

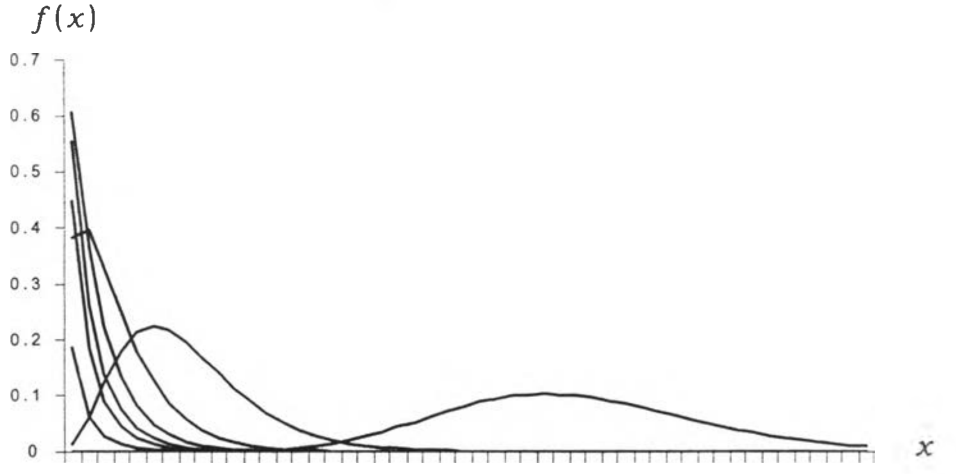
การแจกแจงแกมมาเป็นการแจกแจงแบบต่อเนื่องที่มีประโยชน์ทางวิทยาศาสตร์และเทคโนโลยี ซึ่งการแจกแจงแกมมาเป็นการแจกแจงของตัวแปรสุ่มที่แสดงเวลาที่ต้องรอคอยจนกระทั่งเกิดเหตุการณ์ n เหตุการณ์ เช่น ระยะเวลาที่ต้องใช้ในการบริการลูกค้า n คน เป็นต้น

ให้ X เป็นตัวแปรสุ่มต่อเนื่องที่มีการแจกแจงแกมมา ด้วยพารามิเตอร์ α และ β ดังนั้นฟังก์ชันความหนาแน่นของ X จะอยู่ในรูปของ

$$f(x) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)} x^{\alpha-1} \exp\left(-\frac{x}{\beta}\right) \qquad , x > 0 \quad , \alpha > 0 \quad , \beta > 0$$

เมื่อ สัมประสิทธิ์ความเบ้เท่ากับ $\dfrac{2}{\sqrt{\alpha}}$

และ สัมประสิทธิ์ความโด่งเท่ากับ $3 + \dfrac{6}{\alpha}$



รูปที่ 2.5 แสดงการแจกแจงแกมมา เมื่อ $\beta = 1$ และ

$\alpha = 16, 4, 1.7778, 1, 0.64, 0.4444, 0.16$

โดยค่าน้อยที่สุดมีความโด่งสูงสุดและลดหลั่นกัน ตามลำดับ

<u>คุณสมบัติของการแจกแจงแกมมามีดังนี้</u>

1. ค่าเฉลี่ยและค่าความแปรปรวนจะขึ้นอยู่กับค่าพารามิเตอร์ α และ β ซึ่ง ค่าเฉลี่ย $= \alpha\beta$ และ ค่าความแปรปรวน $= \alpha\beta^2$
2. ลักษณะของโค้งเปลี่ยนไปตามพารามิเตอร์ α
3. สัมประสิทธิ์การแปรผันจะขึ้นอยู่กับค่าพารามิเตอร์ α โดยที่

$$\text{สัมประสิทธิ์การแปรผัน} = \frac{1}{\sqrt{\alpha}}$$

4. การแจกแจงแกมมาที่มีค่า $\alpha = 1$ และ β เป็นค่าใดๆที่มากกว่าศูนย์ จะเป็นการแจกแจงชี้กำลัง (Exponential distribution) ที่มีพารามิเตอร์เป็น β กล่าวคือ $Gamma(1,\beta) = Exp(\beta)$
5. การแจกแจงแกมมาที่มีค่า $\alpha = \frac{k}{2}$ และ $\beta = 2$ จะเป็นการแจกแจงไคกำลังสอง ที่มี ร.ส. เท่ากับ k นั่นคือ $Gamma\left(\frac{k}{2},2\right) = \chi^2_{(k)}$

2.2.4 <u>การแจกแจงไวบูลล์ (Weibull Distribution)</u>

ในปี ค.ศ. 1939 วาลอดดี ไวบูลล์ (Waloddi Weibull) นักฟิสิกส์ชาวสวีเดน เป็นผู้ค้นพบการแจกแจงนี้ขึ้นมา ซึ่งการแจกแจงนี้จะเกี่ยวข้องกับอายุการใช้งานของเครื่องจักรกลต่างๆ

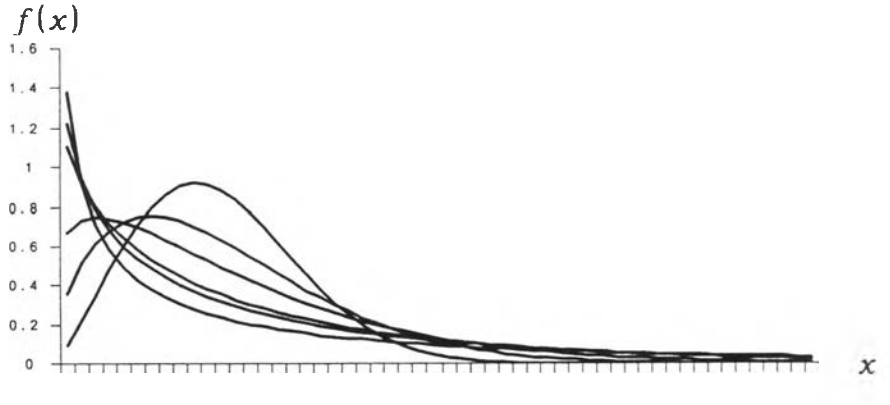
ให้ X เป็นตัวแปรสุ่มต่อเนื่องที่มีการแจกแจงไวบูลล์ ด้วยพารามิเตอร์ α และ β ฟังก์ชันความหนาแน่นของ X จะอยู่ในรูปของ

$$f(x) = \frac{1}{\beta^{\alpha}}\alpha x^{\alpha-1}\exp\left\{-\left(\frac{x}{\beta}\right)^{\alpha}\right\} \qquad ,x>0 \quad ,\alpha>0 \quad ,\beta>0$$

เมื่อ สัมประสิทธิ์ความเบ้เท่ากับ $\dfrac{c-3ab+2a^3}{\left(b-a^2\right)^{3/2}}$

และ สัมประสิทธิ์ความโด่งเท่ากับ $\dfrac{d+6a^2b-3a^4-4ac}{\left(b-a^2\right)^2}$

โดยที่ $a = \Gamma\left(1+\frac{1}{\alpha}\right), b = \Gamma\left(1+\frac{2}{\alpha}\right), c = \Gamma\left(1+\frac{3}{\alpha}\right), d = \Gamma\left(1+\frac{4}{\alpha}\right)$



รูปที่ 2.6 แสดงการแจกแจงไวบูลล์ เมื่อ $\beta = 1$ และ

$\alpha = 2.2156, 1.5639, 1.2111, 0.8631, 0.7686, 0.5737$

โดยค่าน้อยที่สุดมีความโด่งสูงสุดและลดหลั่นกัน ตามลำดับ

<u>คุณสมบัติของการแจกแจงไวบูลล์มีดังนี้</u>

1. ค่าเฉลี่ยและค่าความแปรปรวนจะขึ้นอยู่กับค่าพารามิเตอร์ α และ β โดยที่

$$\text{ค่าเฉลี่ย } = \beta\Gamma\left(1+\frac{1}{\alpha}\right)$$

$$\text{และ ค่าความแปรปรวน } = \beta^2\left\{\Gamma\left(1+\frac{2}{\alpha}\right)-\Gamma^2\left(1+\frac{1}{\alpha}\right)\right\}$$

2. ลักษณะของโค้งเปลี่ยนไปตามพารามิเตอร์ α และจะมีลักษณะการกระจายแบบเบ้ขวา
3. การแจกแจงไวบูลล์ที่มีค่า $\alpha = 1$ และ β เป็นค่าใดๆที่มากกว่าศูนย์ จะมีการแจกแจงชี้กำลัง (Exponential Distribution) ที่มีพารามิเตอร์เป็น β กล่าวคือ $Weibull(1,\beta) = Exp(\beta)$

2.3 การประมาณค่าพารามิเตอร์แบบช่วง

การประมาณค่าแบบช่วง คือ การใช้ช่วงตัวเลขซึ่งคำนวณจากข้อมูลชุดตัวอย่างใดๆมาประมาณค่าพารามิเตอร์ของประชากร ช่วงตัวเลขนี้จะบอกค่าต่ำสุดและค่าสูงสุดที่คาดว่าจะครอบคลุมค่าพารามิเตอร์นั้นๆอยู่ เราเรียกช่วงตัวเลข (L,U) นี้ว่า ช่วงความเชื่อมั่น $(1-\alpha)100$ เปอร์เซนต์ ($(1-\alpha)100\%$ confidence interval) และเรียก $1-\alpha$ ว่า <u>สัมประสิทธิ์ความเชื่อมั่น</u>[1] (confidence coefficient) หรือระดับความเชื่อมั่น (confidence level) โดยเรียก L ว่า ขีดจำกัดล่างของช่วงความเชื่อมั่น (lower confidence limit) และเรียก U ว่า ขีดจำกัดบนของช่วงความเชื่อมั่น (upper confidence limit)

2.4 วิธีการประมาณค่าแบบช่วงที่ใช้ในการวิจัย

2.4.1 <u>วิธีการประมาณค่าแบบช่วงด้วยตัวสถิติที (T)</u>

การประมาณค่าแบบช่วงสำหรับค่าเฉลี่ยของประชากรในกรณีที่ไม่ทราบค่าส่วนเบี่ยงเบนมาตรฐานของประชากรนั้นจะใช้ตัวสถิติทีในการสร้างช่วงความเชื่อมั่น ตัวสถิติทีหรือเรียกอีกอย่างหนึ่งว่าตัวสถิติสติวเดนท์ที (Student's t test) เป็นตัวสถิติที่กอสเสตต์ (Gossette) ได้พัฒนาขึ้น ตัวสถิติทีนี้จะมีการแจกแจงทีด้วย ร.ส. เท่ากับ $n-1$ ซึ่งตัวสถิติทีจะอยู่ในรูปของ

$$t = \frac{\overline{X}-\mu}{S/\sqrt{n}}$$

เมื่อ $\overline{X}$ แทนค่าเฉลี่ยของตัวอย่างเท่ากับ $\dfrac{\sum_{i=1}^{n} X_i}{n}$

---

[1] ความน่าจะเป็นที่ช่วงสุ่มจะครอบคลุมค่าพารามิเตอร์ของประชากร

S แทนส่วนเบี่ยงเบนมาตรฐานของตัวอย่าง[1] เท่ากับ $\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}$

และ n แทนขนาดตัวอย่าง

ดังนั้น ช่วงความเชื่อมั่นบน $(1-\alpha)100\%$ ของค่าเฉลี่ยสำหรับการทดสอบสมมติฐานทางด้านมากกว่า คือ $(0\ ,\ \overline{X} + t_{1-\alpha}(n-1)\frac{S}{\sqrt{n}})$ ส่วนช่วงความเชื่อมั่นล่าง $(1-\alpha)100\%$ ของค่าเฉลี่ยสำหรับการทดสอบสมมติฐานทางด้านน้อยกว่า คือ $(\overline{X} - t_{1-\alpha}(n-1)\frac{S}{\sqrt{n}}\ ,\ \infty)$ และช่วงความเชื่อมั่น $(1-\alpha)100\%$ ของค่าเฉลี่ยสำหรับการทดสอบสมมติฐานสองทาง คือ

$$(\overline{X} - t_{1-\alpha/2}(n-1)\frac{S}{\sqrt{n}}\ ,\ \overline{X} + t_{1-\alpha/2}(n-1)\frac{S}{\sqrt{n}})$$

2.4.2 <u>วิธีการประมาณค่าแบบช่วงด้วยตัวสถิติของจอห์นสัน (J)</u>

ในปี ค.ศ. 1978 จอห์นสัน (Johnson) ได้ทำการแปลงตัวสถิติทีเพื่อให้เหมาะสมสำหรับการประมาณค่าเฉลี่ยของการแจกแจงที่ไม่สมมาตรโดยอาศัยหลักการของ Cornish-Fisher Expansion ซึ่งจะแทนที่ $\overline{X} - \mu$ ในตัวแปรทีด้วยสองเทอมแรกของ inverse Cornish-Fisher Expansion มีรายละเอียดต่างๆดังนี้

รูปแบบทั่วไปของ Cornish-Fisher Expansion สำหรับตัวแปร X คือ

$$CF(X) = \mu + \sigma Z + \left(\frac{\mu_3}{\sigma^2}\right)\left(Z^2 - 1\right) + \cdots$$

เมื่อ Z เป็นตัวแปรสุ่มปกติมาตรฐาน

μ เป็นค่าเฉลี่ยของ X

และ σ^2 , μ_3 เป็นโมเมนต์ศูนย์กลางที่ 2 และ 3 ของ X ตามลำดับ.

ถ้าทุกโมเมนต์หาค่าได้ จะได้ว่า

$$CF(\overline{X}) = \mu + \frac{\sigma}{\sqrt{n}}Z + \frac{\mu_3}{6n\sigma^2}(Z^2 - 1) + O(n^{-3/2})$$

เมื่อ $O(n^{-3/2})$ แทนฟังก์ชันของ n ที่มีระดับของมิติไม่เกิน $n^{-3/2}$

กล่าวคือ $\left|O(n^{-3/2})\right| \le K\,n^{-3/2}$, $K > 0$

---

[1] ในทางทฤษฎี ควรใช้ $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ ซึ่งเป็นตัวประมาณที่ไม่เอนเอียงของ σ^2

ดังนั้น เราจะกำจัดเทอม (Z^2-1) โดยการกำหนด

$$g_1(t)=\left[(\bar{X}-\mu)+\lambda+\delta\left\{(\bar{X}-\mu)^2-\left(\frac{\sigma^2}{n}\right)\right\}\right]\left(\frac{S^2}{n}\right)^{-1/2}$$

ซึ่งสร้างมาจากการแทนที่ $\bar{X}-\mu$ ในตัวเศษของตัวแปรที่ด้วยสองเทอมแรกของ inverse Cornish-Fisher Expansion ของ $\bar{X}-\mu$ โดยที่ λ เป็นฟังก์ชันของ n ที่กำหนดขึ้นเพื่อให้เทอมค่าคงที่มีผลรวมเท่ากับศูนย์ และค่าคงที่ δ กำหนดขึ้นเพื่อให้สัมประสิทธิ์ของเทอม Z^2 ใน Cornish-Fisher Expansion มีค่าเป็นศูนย์ จะได้ว่า $\lambda=\dfrac{\mu_3}{2n\sigma^2}$ และ $\delta=\dfrac{\mu_3}{3\sigma^4}$

ดังนั้น ตัวสถิติของจอห์นสันจะอยู่ในรูปของ

$$g_1(t)=\left\{(\bar{X}-\mu)+\frac{\mu_3}{6\sigma^2 n}+\frac{\mu_3}{3\sigma^4}(\bar{X}-\mu)^2\right\}\left(\frac{S^2}{n}\right)^{-1/2}$$

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

เมื่อ $\hat{\gamma}$ แทนสัมประสิทธิ์ความเบ้เท่ากับ $\dfrac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^3/S^3$

เนื่องจาก $g_1(t)=z$ เป็นสมการกำลังสอง ดังนั้น การหาค่า $g_1^{-1}(z)$ ทำได้โดยใช้สูตรการแก้สมการกำลังสอง ดังนี้

กำหนด $at^2+bt+c=0$ จะได้ว่า $t=\dfrac{-b\pm\sqrt{b^2-4ac}}{2a}$

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

$$\therefore\quad \frac{\hat{\gamma}t^2}{3\sqrt{n}}+t+\left(\frac{\hat{\gamma}}{6\sqrt{n}}-z\right)=0$$

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

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

ดังนั้น ในกรณีที่ $\hat{\gamma}>0$ จะได้เงื่อนไขการผกผันได้ (invertible) ว่า $z\geq\dfrac{\hat{\gamma}}{6\sqrt{n}}-\dfrac{3\sqrt{n}}{4\hat{\gamma}}$

และในกรณีที่ $\hat{\gamma}<0$ จะได้เงื่อนไขการผกผันได้ว่า $z\leq\dfrac{\hat{\gamma}}{6\sqrt{n}}-\dfrac{3\sqrt{n}}{4\hat{\gamma}}$

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

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

เพราะเป็นคำตอบของสมการที่อยู่ในช่วงของฟังก์ชันเพิ่ม (ตัวสถิติของจอห์นสันเป็นการแปลงตัวสถิติทีเพื่อหาค่าเปอร์เซนต์ไทล์ที่ $(1-\alpha)100$ ของการแจกแจงที่สมมาตร ซึ่งค่าเปอร์เซนต์ไทล์นี้มีลักษณะเป็นฟังก์ชันเพิ่ม)

ดังนั้น ช่วงความเชื่อมั่นบน $(1-\alpha)100\%$ ของค่าเฉลี่ยสำหรับการทดสอบสมมติฐานทางด้านมากกว่า คือ $(0\ ,\ \bar{X}-g_1^{-1}(z_\alpha)\frac{S}{\sqrt{n}})$ ส่วนช่วงความเชื่อมั่นล่าง $(1-\alpha)100\%$ ของค่าเฉลี่ยสำหรับการทดสอบสมมติฐานทางด้านน้อยกว่า คือ $(\bar{X}-g_1^{-1}(z_{1-\alpha})\frac{S}{\sqrt{n}}\ ,\ \infty)$ และช่วงความเชื่อมั่น $(1-\alpha)100\%$ ของค่าเฉลี่ยสำหรับการทดสอบสมมติฐานสองทาง คือ

$$(\bar{X}-g_1^{-1}(z_{1-\alpha/2})\frac{S}{\sqrt{n}}\ ,\ \bar{X}-g_1^{-1}(z_{\alpha/2})\frac{S}{\sqrt{n}})$$

2.4.3 <u>วิธีการประมาณค่าแบบช่วงด้วยตัวสถิติของฮอลล์ (H)</u>

ในปี ค.ศ. 1992 ฮอลล์ (Hall) ได้ทำการแปลงตัวสถิติทีเพื่อให้มีความเหมาะสมสำหรับการประมาณค่าเฉลี่ยของการแจกแจงที่ไม่สมมาตรเช่นเดียวกันแต่อาศัยหลักการของ Edgeworth Expansion มีรายละเอียดต่างๆดังนี้

กำหนดให้ Edgeworth Expansion มีรูปแบบคือ

$$P(T \le x) = \Phi(x) + n^{-1/2}\gamma(ax^2+b)\phi(x) + O(n^{-1})$$

เมื่อ a และ b เป็นค่าคงที่ที่ทราบค่า

γ เป็นค่าคงที่ที่ประมาณได้

Φ และ ϕ เป็นฟังก์ชันการแจกแจงสะสมและฟังก์ชันความหนาแน่นของการแจกแจงแบบปกติมาตรฐานตามลำดับ

กำหนดให้ $$T_1 = T_1(K) = K + a\hat{\gamma}K^2 + \frac{1}{3}a^2\hat{\gamma}^2K^3 + n^{-1}b\hat{\gamma} \qquad (2.1)$$

เมื่อ $\hat{\gamma}$ เป็นตัวประมาณของ γ

กำหนดให้ $X_1, X_2, \ldots, X_n$ เป็นค่าสังเกตที่เป็นอิสระซึ่งกันและกันจากการแจกแจงที่มีค่าเฉลี่ย μ และความแปรปรวน σ^2

$$\therefore \quad K = \frac{(\hat{\theta}-\theta)}{\hat{\sigma}} = \frac{(\bar{X}-\mu)}{S} = \frac{t}{\sqrt{n}}, a = \frac{1}{3}, b = \frac{1}{6}$$

แทนค่าในสมการที่ (2.1) จะได้

$$T_1 = \frac{t}{\sqrt{n}} + \frac{1}{3n}\hat{\gamma}t^2 + \frac{1}{27}\hat{\gamma}^2 n^{-3/2}t^3 + \frac{1}{6n}\hat{\gamma}$$

ดังนั้น ตัวสถิติของฮอลล์จะอยู่ในรูปของ

$$g_2(t) = T_1\sqrt{n} = t + \frac{1}{3\sqrt{n}}\hat{\gamma}t^2 + \frac{1}{27n}\hat{\gamma}^2t^3 + \frac{1}{6\sqrt{n}}\hat{\gamma}$$

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

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

ดังนั้น ช่วงความเชื่อมั่นบน $(1-\alpha)100\%$ ของค่าเฉลี่ยสำหรับการทดสอบสมมติฐานทางด้านมากกว่า คือ $(0\ ,\ \bar{X} - g_2^{-1}(z_\alpha)\frac{S}{\sqrt{n}})$ ส่วนช่วงความเชื่อมั่นล่าง $(1-\alpha)100\%$ ของค่าเฉลี่ยสำหรับการทดสอบสมมติฐานทางด้านน้อยกว่า คือ $(\bar{X} - g_2^{-1}(z_{1-\alpha})\frac{S}{\sqrt{n}}\ ,\ \infty)$ และช่วงความเชื่อมั่น $(1-\alpha)100\%$ ของค่าเฉลี่ยสำหรับการทดสอบสมมติฐานสองทาง คือ

$$(\bar{X} - g_2^{-1}(z_{1-\alpha/2})\frac{S}{\sqrt{n}}\ ,\ \bar{X} - g_2^{-1}(z_{\alpha/2})\frac{S}{\sqrt{n}})$$

2.4.4 <u>วิธีการประมาณค่าแบบช่วงด้วยตัวสถิติของเชน (C)</u>

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

จาก Edworth Expansion จะได้ว่า

$$P\left\{\frac{\sqrt{n}(\bar{X}-\mu)}{S} \le x - \frac{1}{6\sqrt{n}}\hat{\gamma}(1+2x^2)\right\} = \mathrm{F}(x) + O(n^{-1/2}) \qquad (2.2)$$

ในการหาตัวประมาณที่มีความแม่นยำจะต้องแก้อสมการ

$$\left\{\frac{\sqrt{n}(\bar{X}-\mu)}{S} > x - \frac{1}{6\sqrt{n}}\hat{\gamma}(1+2x^2)\right\}$$

กำหนดให้ $a = \frac{\hat{\gamma}}{6\sqrt{n}}$ และ $t = \frac{\sqrt{n}(\bar{X}-\mu)}{S}$

เมื่อ n มีค่ามาก จะได้ว่า $1 - 8a(t+a) \ge 0$

จากสมการที่ (2.2) จะได้ว่า

$$P\left\{\frac{1-\sqrt{1-8a(t+a)}}{4a} > z_\alpha \quad \text{หรือ} \quad \frac{1+\sqrt{1-8a(t+a)}}{4a} < z_\alpha\right\} = \alpha + O(n^{-1/2})$$

ดังนั้น

$$P\left\{\frac{1-\sqrt{1-8a(t+a)}}{4a} > z_{\alpha}\right\} \le \alpha$$

จาก Taylor Expansion จะได้ว่า

$$\frac{1-\sqrt{1-8a(t+a)}}{4a} = t+a+2at^2+4a^2(t+2t^3)+O(n^{-1})$$

ดังนั้น ตัวสถิติของเชนจะอยู่ในรูปของ

$$g_3(t) = t+a(1+2t^2)+4a^2(t+2t^3)$$

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

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

กำหนดให้ $f(t) = g_3(t) - z = 0$

$$\therefore \quad f(t) = t+\frac{\hat{\gamma}}{\sqrt{n}}\left(\frac{t^2}{3}+\frac{1}{6}\right)+\frac{\hat{\gamma}^2}{9n}\left(t+2t^3\right)-z$$

1. กำหนดค่าเริ่มต้น คือ t_0 และค่าความคลาดเคลื่อนที่ยอมรับได้ (tolerence)
2. คำนวณค่า $t_{n+1} = t_n - \dfrac{f(t_n)}{f'(t_n)}$ สำหรับ $n = 0,1,2,\ldots$

   เมื่อ $f'(t)$ คือ ค่าอนุพันธ์อันดับที่หนึ่งของ f ที่จุด t

   $$\therefore \quad f'(t) = 1+\frac{2\hat{\gamma}t}{3\sqrt{n}}+\frac{\hat{\gamma}^2}{9n}+\frac{2\hat{\gamma}^2t^2}{3n} = 1+\frac{2\hat{\gamma}t}{3\sqrt{n}}+\frac{\hat{\gamma}^2}{3n}\left(\frac{1}{3}+2t^2\right)$$
3. คำนวณซ้ำในข้อ 2 จนกว่า $\left|\dfrac{f(t_{n+1})}{f'(t_{n+1})}\right| < tolerence$ จะได้ว่าคำตอบของสมการ

   คือ $g_3^{-1}(z) = t_{n+1}$

ดังนั้น ช่วงความเชื่อมั่นบน $(1-\alpha)100\%$ ของค่าเฉลี่ยสำหรับการทดสอบสมมติฐานทางด้านมากกว่า คือ $(0\,,\ \bar{X}-g_3^{-1}(z_{\alpha})\frac{S}{\sqrt{n}})$ ส่วนช่วงความเชื่อมั่นล่าง $(1-\alpha)100\%$ ของค่าเฉลี่ยสำหรับการทดสอบสมมติฐานทางด้านน้อยกว่า คือ $(\bar{X}-g_3^{-1}(z_{1-\alpha})\frac{S}{\sqrt{n}}\,,\ \infty)$ และช่วงความเชื่อมั่น $(1-\alpha)100\%$ 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 การประมาณค่าแบบช่วงจะนำวิธีการบูตสแตรป (Bootstrap method) มาช่วยในการหาค่าเปอร์เซนต์ไทล์ที่ $\alpha \times 100$ ของการแจกแจงของ $g(t)$ สำหรับการสร้างช่วงความเชื่อมั่น โดยกำหนดให้ $X^* = (X_1^*, \ldots, X_n^*)'$ แทนตัวอย่างสุ่มขนาด n ซึ่งสุ่มแบบใส่คืนจากตัวอย่างชุดแรก $X = (X_1, \ldots, X_n)'$ 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[1]

1. ถ้า $B\alpha$ เป็นจำนวนเต็มแล้ว $\hat{t}^{(\alpha)}$ จะเป็นค่าที่ $B\alpha$ ของ $g^*(t^*)_b$
2. ถ้า $B\alpha$ ไม่เป็นจำนวนเต็มแล้ว $\hat{t}^{(\alpha)}$ จะเป็นค่าที่ k ของ $g^*(t^*)_b$ เมื่อ k คือจำนวนเต็มที่มากที่สุดที่น้อยกว่าหรือเท่ากับ $(B+1)\alpha$

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

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

2.5 เกณฑ์ในการตัดสินใจ

ในการเปรียบเทียบวิธีการประมาณค่าแบบช่วงสำหรับค่าเฉลี่ยของประชากรที่มีการแจกแจงแบบเบ้ขวาจะพิจารณาเปรียบเทียบค่าสัมประสิทธิ์ความเชื่อมั่น ค่าเฉลี่ยของขีดจำกัดความเชื่อมั่นล่าง (MLCL) ค่าเฉลี่ยของขีดจำกัดความเชื่อมั่นบน (MUCL) และค่าเฉลี่ยความยาวของช่วงความเชื่อมั่นของแต่ละวิธีการประมาณ (MCIL) โดยในขั้นแรกจะต้องตรวจสอบว่าช่วงความเชื่อมั่นที่ได้จากทั้ง 4 วิธีนั้นให้ค่าสัมประสิทธิ์ความเชื่อมั่นจากการทดลองเท่ากับค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนดหรือไม่ แล้วจึงพิจารณาค่าเฉลี่ยของขีดจำกัดความเชื่อมั่นล่างจากการทดสอบสมมติฐานทางด้านน้อยกว่า ค่าเฉลี่ยของขีดจำกัดความเชื่อมั่นบนจากการทดสอบสมมติฐานทางด้านมากกว่า และค่าเฉลี่ยความยาวของช่วงความเชื่อมั่นจากการทดสอบสมมติฐานสองทาง สำหรับการพิจารณาจะพิจารณาว่าค่าเฉลี่ยของขีดจำกัดความเชื่อมั่นล่างของวิธีการใดให้ค่าสูงที่สุดและค่าเฉลี่ยของขีดจำกัดความเชื่อมั่นบนของวิธีการใดให้ค่าต่ำที่สุด ส่วนค่าเฉลี่ยความ

---

[1] เกณฑ์ที่กำหนดเพื่อความสะดวกในการคำนวณ แต่ทางทฤษฎีควรพิจารณา $\hat{t}^{(\alpha)}$ เป็นค่าที่ $(B+1)\alpha$ ของ $g^*(t^*)_b$

ยาวของช่วงความเชื่อมั่นจากการทดสอบสมมติฐานสองทางจะพิจารณาว่าค่าเฉลี่ยความยาวของช่วงความเชื่อมั่นของวิธีการใดให้ค่าต่ำที่สุดจะเป็นวิธีการที่เหมาะสมสำหรับการประมาณค่าแบบช่วงภายใต้การทดสอบสมมติฐานนั้นๆ ทั้งนี้จะพิจารณาเฉพาะกรณีที่วิธีการประมาณนั้นให้ค่าสัมประสิทธิ์ความเชื่อมั่นจากการทดลองไม่ต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนดเท่านั้น

2.5.1 การตรวจสอบค่าสัมประสิทธิ์ความเชื่อมั่น

การคำนวณค่าสัมประสิทธิ์ความเชื่อมั่นจะทำการพิจารณาว่าช่วงความเชื่อมั่นที่ได้จากแต่ละวิธีการประมาณนั้นครอบคลุมค่าพารามิเตอร์หรือไม่ โดยทำการนับจำนวนครั้งทั้งหมดที่ช่วงความเชื่อมั่นที่ประมาณได้ครอบคลุมค่าพารามิเตอร์ นำค่าที่ได้นี้หารด้วยจำนวนครั้งที่ทำซ้ำทั้งหมด ค่าที่ได้นี้คือค่าสัมประสิทธิ์ความเชื่อมั่นที่ได้จากการทดลองของแต่ละวิธีการประมาณ ในการตรวจสอบว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่ได้จากการทดลองของวิธีการใดมีค่าไม่ต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนดนั้นจะอาศัยการทดสอบสมมติฐานด้วยตัวสถิติ Z มีรายละเอียดดังนี้

กำหนดให้ X เป็นจำนวนครั้งทั้งหมดที่ช่วงความเชื่อมั่นที่ประมาณได้ครอบคลุมค่าพารามิเตอร์ในการทดลองแบบ $Ber(p)$ จำนวน n ครั้งซึ่งเป็นอิสระซึ่งกันและกัน โดยที่ p เป็นความน่าจะเป็นที่ช่วงความเชื่อมั่นที่ประมาณได้จะครอบคลุมค่าพารามิเตอร์ เมื่อจำนวนครั้งของการทดลองมีค่ามาก X จะมีการแจกแจงโดยประมาณเป็น $N(np,npq)$ และกำหนด $\hat{P} = X/n$ เป็นค่าสัมประสิทธิ์ความเชื่อมั่นจากการทดลอง โดยที่ $\hat{P}$ จะมีการแจกแจงโดยประมาณเป็น $N(p,\frac{pq}{n})$ และ $(0 \le p \le 1)$

ดังนั้นในการพิจารณาว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่ได้จากการทดลองของวิธีการใดมีค่าไม่ต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนดนั้นจะใช้การทดสอบสมมติฐานด้วยตัวสถิติ Z เพื่อให้อำนาจของการทดสอบ (power of test) มีค่ามาก สมมติฐานที่ใช้ในการทดสอบ คือ

$H_0 : p \le p^*$

$H_1 : p > p^*$

เราจะปฏิเสธสมมติฐาน H_0 เมื่อ $\frac{\hat{P}-p}{\sqrt{p(1-p)/n}} > z_{1-\alpha}$

$$\hat{P} > p + z_{1-\alpha}\sqrt{p(1-p)/3000}$$

1. ค่าสัมประสิทธิ์ความเชื่อมั่น 0.90

$H_0 : p \le 0.90$

$H_1 : p > 0.90$

จะได้ว่า เราจะปฏิเสธสมมติฐาน H_0 เมื่อ $\hat{P} > 0.90 + 1.282\sqrt{(0.90 \times 0.10)/3000}$

$$\hat{P} > 0.9070$$

เพราะฉะนั้น ถ้าวิธีการใดให้ค่าสัมประสิทธิ์ความเชื่อมั่นจากการทดลองไม่ต่ำกว่า 0.9070 แสดงว่าวิธีการนั้นให้ค่าสัมประสิทธิ์ความเชื่อมั่นตามค่าที่กำหนด

2. ค่าสัมประสิทธิ์ความเชื่อมั่น 0.95

$H_0 : p \le 0.95$

$H_1 : p > 0.95$

จะได้ว่า เราจะปฏิเสธสมมติฐาน H_0 เมื่อ $\hat{P} > 0.95 + 1.645\sqrt{(0.95 \times 0.05)/3000}$

$$\hat{P} > 0.9565$$

เพราะฉะนั้น ถ้าวิธีการใดให้ค่าสัมประสิทธิ์ความเชื่อมั่นจากการทดลองไม่ต่ำกว่า 0.9565 แสดงว่าวิธีการนั้นให้ค่าสัมประสิทธิ์ความเชื่อมั่นตามค่าที่กำหนด

3. ค่าสัมประสิทธิ์ความเชื่อมั่น 0.99

$H_0 : p \le 0.99$

$H_1 : p > 0.99$

จะได้ว่า เราจะปฏิเสธสมมติฐาน H_0 เมื่อ $\hat{P} > 0.99 + 2.326\sqrt{(0.99 \times 0.01)/3000}$

$$\hat{P} > 0.9942$$

เพราะฉะนั้น ถ้าวิธีการใดให้ค่าสัมประสิทธิ์ความเชื่อมั่นจากการทดลองไม่ต่ำกว่า 0.9942 แสดงว่าวิธีการนั้นให้ค่าสัมประสิทธิ์ความเชื่อมั่นตามค่าที่กำหนด

2.5.2 <u>การหาค่าเฉลี่ยของขีดจำกัดความเชื่อมั่นล่าง ค่าเฉลี่ยของขีดจำกัดความเชื่อมั่นบนและค่าเฉลี่ยความยาวของช่วงความเชื่อมั่น</u>

การคำนวณค่าเฉลี่ยของขีดจำกัดความเชื่อมั่นล่างจากการทดสอบสมมติฐานทางด้านน้อยกว่า (MLCL) ทำได้โดยหาค่าผลบวกสะสมของขีดจำกัดความเชื่อมั่นล่างแล้วหารด้วยจำนวนครั้งที่ทำซ้ำทั้งหมด มีรูปแบบการคำนวณดังนี้

$$\text{MLCL} = \frac{\sum_{i=1}^{3000} L_i}{3000}$$

ในทำนองเดียวกัน การคำนวณค่าเฉลี่ยของขีดจำกัดความเชื่อมั่นบนจากการทดสอบสมมติฐานทางด้านมากกว่า (MUCL) ทำได้โดยหาค่าผลบวกสะสมของขีดจำกัดความเชื่อมั่นบนแล้วหารด้วยจำนวนครั้งที่ทำซ้ำทั้งหมด มีรูปแบบการคำนวณดังนี้

$$\text{MUCL} = \frac{\sum_{i=1}^{3000} U_i}{3000}$$

ส่วนค่าเฉลี่ยความยาวของช่วงความเชื่อมั่นจากการทดสอบสมมติฐานสองทาง (MCIL) จะคำนวณจากผลบวกสะสมของผลต่างระหว่างขีดจำกัดความเชื่อมั่นบนและขีดจำกัดความเชื่อมั่นล่างหารด้วยจำนวนครั้งที่ทำซ้ำทั้งหมด มีรูปแบบการคำนวณดังนี้

$$\text{MCIL} = \frac{\sum_{i=1}^{3000} (U_i - L_i)}{3000}$$

2.6 ผลงานวิจัยที่เกี่ยวข้อง

สำหรับผลงานวิจัยที่เกี่ยวข้องกับการสร้างช่วงความเชื่อมั่นสำหรับค่าเฉลี่ยของประชากรที่มีการแจกแจงแบบเบ้ขวามีแนวความคิดมาจากการหาตัวประมาณที่เหมาะสมสำหรับการประมาณค่าเฉี่ยของประชากรที่ไม่มีการแจกแจงเป็นแบบปกติ ซึ่งมีรายละเอียดดังนี้

ในปี ค.ศ. 1928 ซอร์ฟิซเตอร์ (Sophister) , เนย์แมน (Neyman) และเพียร์สัน (Pearson) ได้แสดงให้เห็นถึงผลกระทบจากการใช้ตัวสถิติที เมื่อประชากรไม่ได้มีการแจกแจงเป็นแบบปกติและขนาดตัวอย่างมีขนาดเล็ก และเขาพบว่าความเบ้ (skewness) ของประชากรจะมีผลต่อการแจกแจงของตัวสถิติทีมากกว่าความโด่ง (kurtosis) และประชากรที่มีลักษณะเบ้ขวา (positively skewed distribution) มีผลกระทบต่อการแจกแจงของตัวสถิติทีในทางเบ้ซ้าย (negatively skewed distribution) ในทางกลับกัน ประชากรที่มีลักษณะเบ้ซ้ายจะมีผลกระทบต่อการแจกแจงของตัวสถิติทีในทางเบ้ขวา ดังนั้นจึงได้เสนอตัวสถิติ $|t|$ ในการสร้างช่วงความเชื่อมั่นสำหรับค่าเฉลี่ยเพื่อลดผลกระทบจากความเบ้ แต่การใช้ตัวสถิตินี้จะให้ผลถูกต้องเมื่อประชากรมีความเบ้เพียงเล็กน้อยเท่านั้น ในปี ค.ศ. 1935 บาร์ทเลต (Bartlett) ค.ศ. 1936 เกียร์รี่ (Geary) และ ค.ศ. 1949 เกย์เยน (Gayen) ได้ศึกษาการแจกแจงทีโดยใช้ค่าเฉลี่ยของ Edgeworth หรือ Gram-Charlier Expansion การประมาณค่าพารามิเตอร์ของการแจกแจงของบาร์ทเลตและเกียร์รี่จะมีความคล้ายคลึงกับเกย์เยน ซึ่งในทางปฏิบัติแล้วจะยุ่งยาก ดังนั้น เกย์เยนจึงได้สมมติให้ทราบค่าพารามิเตอร์ของประชากร ซึ่งไม่เป็นจริงในทางปฏิบัติ ต่อมาในปี ค.ศ. 1955 บอกซ์ (Box) และ แอนเดอร์สัน (Anderson) ได้เสนอเทคนิคนอนพาราเมตริก (Nonparametric) มาใช้ในการแก้ปัญหาที่เกิดจากการแจกแจงของประชากร

ในปี ค.ศ. 1961 โฮเทลลิ่ง (Hotelling) ได้ศึกษาค้นพบการกระจายที่อธิบายอัตราส่วนของพื้นที่หางของการแจกแจงทีสำหรับตัวอย่างจากการแจกแจงคงที่ เขาพบว่าถ้าการแจกแจงของประชากรมีลักษณะหางใหญ่จะมีผลทำให้การแจกแจงทีมีลักษณะหางเล็ก ในปี ค.ศ. 1978 จอห์นสัน (Johnson) ได้เสนอตัวสถิติ modified t-test ซึ่งได้ทำการแปลงตัวสถิติทีเพื่อให้มี

ความเหมาะสมสำหรับการประมาณค่าเฉลี่ยของการแจกแจงที่ไม่สมมาตรโดยใช้ Cornish-Fisher Expansion ตัวสถิติ modified t-test นี้จะลดความเอนเอียงของตัวสถิติทีลงและมีการแจกแจงที่ไม่ขึ้นอยู่กับประชากรที่มีการแจกแจงไม่ปกติ แต่วิธีการของจอห์นสันนี้ยังคงไม่เหมาะสมในกรณีที่ขนาดตัวอย่างมีขนาดเล็กและประชากรมีความเบ้มาก ต่อมาในปี ค.ศ. 1992 ฮอลล์ (Hall) ได้เสนอวิธีการที่ปรับปรุงแก้ไขเพิ่มความถูกต้องจากวิธีการของจอห์นสันโดยทำการแปลงตัวสถิติทีด้วย Edgeworth Expansion เขาพบว่าช่วงความเชื่อมั่นด้านล่างที่ได้ให้ความแม่นยำสูงสำหรับค่าเฉลี่ยของประชากรที่ไม่สมมาตรต่างๆ แต่ช่วงความเชื่อมั่นด้านบนยังให้ความแม่นยำไม่ค่อยดีนักสำหรับตัวอย่างขนาดเล็กและเขาพบว่าในบางกรณีตัวสถิติของจอห์นสันจะให้ขีดจำกัดความเชื่อมั่นล่างสูงกว่าขีดจำกัดความเชื่อมั่นบน

ในปี ค.ศ. 1993 ซัตตัน (Sutton) ได้เสนอวิธีการสุ่มตัวอย่างซ้ำของบูตสแตรป (Bootstrap resampling) มาใช้ในการประมาณค่าเฉลี่ยของประชากรที่ไม่สมมาตร ซึ่งวิธีการนี้จะให้ความแม่นยำมากกว่าตัวสถิติของจอห์นสัน ในปี ค.ศ. 1995 เชน (Chen) ได้เสนอตัวสถิติทดสอบใหม่สำหรับค่าเฉลี่ยของประชากรที่มีการแจกแจงแบบเบ้ขวาด้วยกลุ่มตัวอย่างขนาดเล็ก ซึ่งทำการแปลงตัวสถิติทีโดยใช้ Edgeworth Expansion และ Taylor Expansion เขาพบว่าตัวสถิติของเขามีความแม่นยำในกรณีตัวอย่างขนาดเล็ก นอกจากนี้ ในปี ค.ศ. 2000 ซู (Zhou) และ เกาว์ (Gao) ได้แสดงให้เห็นว่าการนำวิธีการบูตสแตรป (Bootstrap method) มาช่วยในการหาค่าของตัวสถิติทีที่แปลงจาตัวสถิติทีทั้งของจอห์นสันและฮอลล์จะทำให้ผลที่ได้มีความแม่นยำมากยิ่งขึ้น