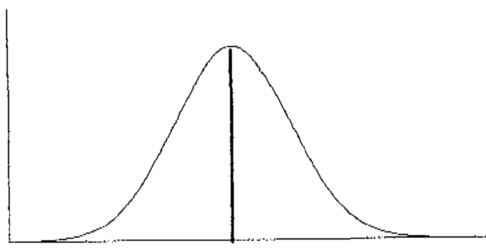
บทที่ 2

ทฤษฎีและผลงานวิจัยที่เกี่ยวข้อง

2.1 ทฤษฎีที่เกี่ยวข้อง

2.1.1 ความเบ้ (Skewness)

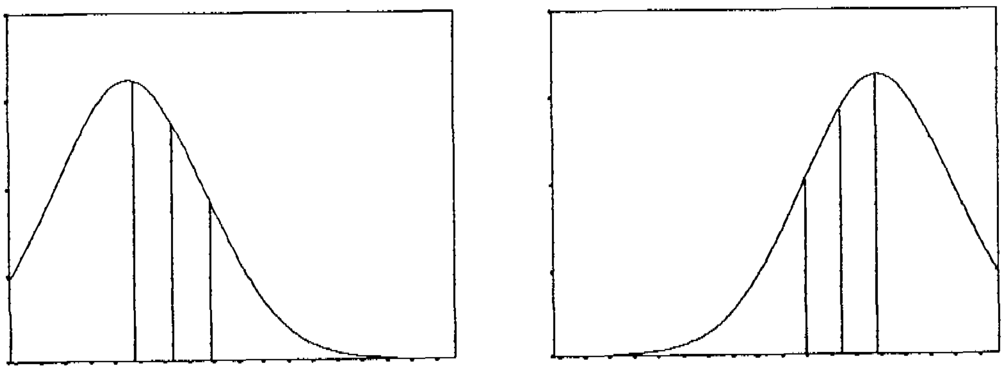
ประชากรที่มีการแจกแจงสมมาตร เส้นโค้งที่ได้จะมีลักษณะเป็นรูประฆังคว่ำที่สมมาตรรอบค่าเฉลี่ย โดยเส้นโค้งทางด้านขวามือของค่าเฉลี่ยและทางด้านซ้ายของค่าเฉลี่ยจะมีลักษณะเหมือนกันทุกประการคือ เป็นข้อมูลที่เบี่ยงเบนจากค่ากลางไปในทางบวกและทางลบพอๆ กัน หรือเป็นข้อมูลที่มีการกระจายสม่ำเสมอ ดังนั้นค่าเฉลี่ย, มัธยฐาน, และฐานนิยม จะมีค่าเท่ากันหรือทับกันพอดี



ค่าเฉลี่ย = ค่ามัธยฐาน = ค่าฐานนิยม

รูปที่ 2.1 แสดงเส้นโค้งความถี่ของข้อมูลที่มีการแจกแจงสมมาตร

สำหรับประชากรที่มีการแจกแจงไม่สมมาตร เส้นโค้งความถี่ของข้อมูลจะมีลักษณะเบ้ไปข้างใดข้างหนึ่ง ถ้าค่าเฉลี่ยมีค่ามากกว่ามัธยฐานและมัธยฐานมีค่ามากกว่าฐานนิยม เส้นโค้งที่ได้จะมีลักษณะเบ้ขวา (Skewed to Right) และมีค่าความเบ้เป็นบวก (Positively Skewed) ดังรูป 2.2 (ก) ถ้าฐานนิยมมีค่ามากกว่ามัธยฐานและมัธยฐานมีค่ามากกว่าค่าเฉลี่ย เส้นโค้งที่ได้จะมีลักษณะเบ้ซ้าย (Skewed to Left) และมีค่าความเบ้เป็นลบ (Negatively Skewed) ดังรูป 2.2 (ข)



ค่าฐานนิยม < ค่ามัธยฐาน< ค่าเฉลี่ย

(ก) เบ้ขวา

ค่าเฉลี่ย < ค่ามัธยฐาน< ค่าฐานนิยม

(ข) เบ้ซ้าย

รูปที่ 2.2 แสดงเส้นโค้งความถี่ของข้อมูลที่มีการแจกแจงไม่สมมาตร

การวัดความเบ้ (Measure of Skewness) จะใช้สัมประสิทธิ์ความเบ้เป็นเครื่องมือวัด โดยในที่นี้จะหาค่าสัมประสิทธิ์ความเบ้ (Coefficient of Skewness) โดยวิธีโมเมนต์ (Classical Moment / Conventional Moment) โดยมีสูตรสัมประสิทธิ์ความเบ้ประชากร (Bowley, 1901) คือ

$$\gamma_1 = \frac{E(X-\mu)^3}{\left(\sqrt{E(X-\mu)^2}\right)^3} = \frac{\mu_3}{\left(\sqrt{\mu_2}\right)^3} = \frac{\mu_3}{\sigma^3}$$

โดยที่ γ_1 คือ สัมประสิทธิ์ความเบ้ของข้อมูลประชากร

μ_3 คือ โมเมนต์ศูนย์กลางอันดับที่ 3 เท่ากับ $E(X-\mu)^3$

μ_2 คือ โมเมนต์ศูนย์กลางอันดับที่ 2 เท่ากับ $E(X-\mu)^2 = Var(X) = \sigma^2$

การวัดความเบ้ด้วยวิธีโมเมนต์ จะให้ค่าต่างๆกันดังนี้

- กรณี $\gamma_1 = 0$ ประชากรมีการแจกแจงสมมาตร
- กรณี $\gamma_1 > 0$ ประชากรมีการแจกแจงเบ้ขวา
- กรณี $\gamma_1 < 0$ ประชากรมีการแจกแจงเบ้ซ้าย

2.1.2 ความโด่ง (Kurtosis)

การวัดความโด่ง คือ การวัดเส้นโค้งว่าจะมีความโด่งมากน้อยเพียงใด โดยเส้นโค้งที่ได้จากการแจกแจงแบบปกติจะเรียกว่า เส้นโค้งปกติ ส่วนเส้นโค้งใดที่โด่งผิดจากเส้นโค้งปกติก็นับเป็นเส้นโค้งที่ไม่ปกติ แม้แต่เป็นรูประฆังที่สมมาตรก็ตาม การวัดความโด่ง (Measure of Kurtosis) จะใช้สัมประสิทธิ์ความโด่งเป็นเครื่องมือวัด โดยในที่นี้จะหาค่าสัมประสิทธิ์ความโด่ง (Coefficient of Kurtosis) โดยวิธีโมเมนต์ (Classical Moment / Conventional Moment) โดยมีสูตรสัมประสิทธิ์ความโด่งประชากร (Bowley, 1901) คือ

$$\gamma_2 = \frac{E(X-\mu)^4}{\left(\sqrt{E(X-\mu)^2}\right)^4} = \frac{\mu_4}{\left(\sqrt{\mu_2}\right)^4} = \frac{\mu_4}{\sigma^4}$$

โดยที่ γ_2 คือ สัมประสิทธิ์ความโด่งของข้อมูลประชากร

μ_4 คือ โมเมนต์ศูนย์กลางอันดับที่ 4 เท่ากับ $E(X-\mu)^4$

μ_2 คือ โมเมนต์ศูนย์กลางอันดับที่ 2 เท่ากับ $E(X-\mu)^2 = Var(X) = \sigma^2$

การวัดความโด่งด้วยวิธีโมเมนต์ จะให้ค่าต่างๆกันดังนี้

- กรณี $\gamma_2 = 3$ แสดงว่าเส้นโค้งมีความโด่งเป็นปกติ จะเรียกว่า Mesokurtic
- กรณี $\gamma_2 < 3$ แสดงว่าเส้นโค้งมีลักษณะแบนราบกว่าปกติ จะเรียกว่า Platykurtic
- กรณี $\gamma_2 > 3$ แสดงว่าเส้นโค้งมีลักษณะสูงกว่าปกติ จะเรียกว่า Leptokurtic

2.2 การแจกแจงที่ใช้ในการวิจัย

2.2.1 การแจกแจงแบบปกติ (Normal Distribution)

กำหนดให้ X เป็นตัวแปรสุ่มต่อเนื่องที่มีการแจกแจงแบบปกติ ด้วยพารามิเตอร์ μ และ σ^2 จะได้ฟังก์ชันการแจงแจง (Distribution function) หรือ ฟังก์ชันความหนาแน่นความน่าจะเป็น (Probability density function) ของ X คือ

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

โดยที่ μ คือ ค่าเฉลี่ย

σ^2 คือ ความแปรปรวน

ลักษณะทั่วไปของตัวแปรสุ่ม X ที่มีการแจกแจงแบบปกติ

1. ค่าเฉลี่ยของตัวแปรสุ่ม X

   $E(X) = \mu$

2. ความแปรปรวนของตัวแปรสุ่ม X

   $Var(X) = \sigma^2$

3. สัมประสิทธิ์ความเบ้

   $\gamma_1 = 0$

4. สัมประสิทธิ์ความโด่ง

   $\gamma_2 = 3$

และการแจกแจงแบบปกติมีคุณสมบัติดังนี้

1. ฟังก์ชันความหนาแน่นความน่าจะเป็นมีลักษณะเป็นเส้นโค้งไม่มีความเบ้ หรือความเบ้เท่ากับศูนย์
2. เส้นโค้งปกติเป็นรูประฆังคว่ำสมมาตรรอบ $X = \mu$
3. ค่าเฉลี่ย มัธยฐาน และฐานนิยมมีค่าเท่ากันและอยู่ในตำแหน่งกลางเส้นโค้งในแนวตั้งฉาก
4. มีแกน X เป็นเส้นกำกับ (Asymptote) กล่าวคือปลายของเส้นโค้งทั้งสองข้างจะไม่ตัดแกน X

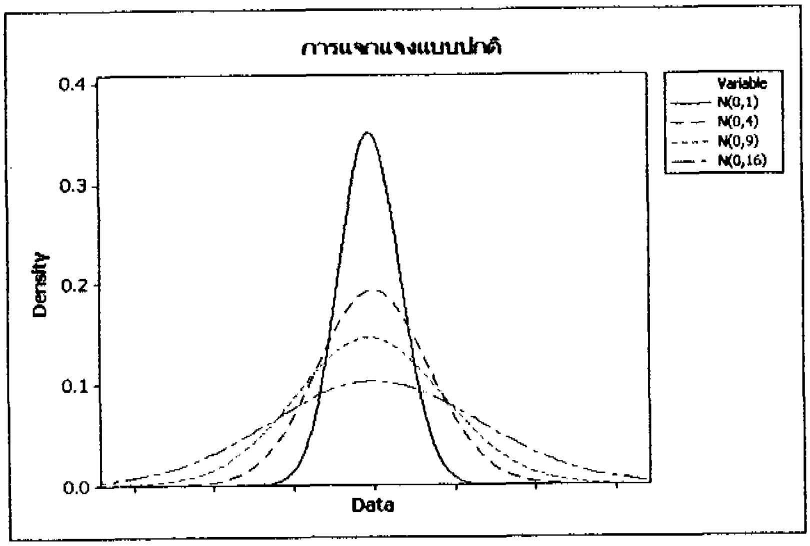
5. พื้นที่ใต้โค้งทั้งหมดคือความน่าจะเป็นของแซมเปิลสเปซ (Sample space) ซึ่งมีค่าเท่ากับ 1 หรือ 100%

6. พื้นที่ใต้โค้งระหว่างจุดที่เป็น $\pm1\sigma$ มีประมาณ 68.26%, พื้นที่ใต้โค้งระหว่างจุดที่เป็น $\pm2\sigma$ มีประมาณ 95.46% และพื้นที่ใต้โค้งระหว่างจุดที่เป็น $\pm3\sigma$ มีประมาณ 99.74%

7. เมื่อการแจกแจงปกติมีค่า μ เท่ากับ 0 และ σ เท่ากับ 1 จะมีการแจกแจงที่เรียกว่าการแจกแจงปกติมาตรฐาน (Standard normal distribution)

รูปที่ 2.3

แสดงฟังก์ชันความหนาแน่นของการแจกแจงแบบปกติ



2.2.2 การแจกแจงแบบปกติมาตรฐาน (Standard normal distribution)

การแจกแจงแบบปกติมาตรฐานเป็นรูปแบบการแจกแจงหนึ่งของการแจกแจงแบบปกติ ในกรณีที่มีค่าเฉลี่ย (μ) เท่ากับ 0 และความแปรปรวน (σ^2) เท่ากับ 1

กำหนดให้ X เป็นตัวแปรสุ่มต่อเนื่องที่มีการแจกแจงแบบปกติ ด้วยพารามิเตอร์ μ=0 และ σ^2=1 จะได้ฟังก์ชันการแจงแจง (Distribution function) หรือ ฟังก์ชันความหนาแน่นความน่าจะเป็น (Probability density function) ของ X คือ

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

โดยที่ μ คือ ค่าเฉลี่ย

σ^2 คือ ความแปรปรวน

ลักษณะทั่วไปของตัวแปรสุ่ม X ที่มีการแจกแจงแบบปกติมาตรฐาน

1. ค่าเฉลี่ยของตัวแปรสุ่ม X

   $E(X) = \mu = 0$

2. ความแปรปรวนของตัวแปรสุ่ม X

   $Var(X) = \sigma^2 = 1$

3. สัมประสิทธิ์ความเบ้

   $\gamma_1 = 0$

4. สัมประสิทธิ์ความโด่ง

   $\gamma_2 = 3$

2.2.3 การแจกแจงแบบที (t Distribution)

กำหนดให้ Z เป็นตัวแปรสุ่มที่มีการแจกแจงแบบปกติมาตรฐาน และ Y เป็นตัวแปรสุ่มที่มีการแจกแจงแบบไคกำลังสองด้วยระดับขั้นความเสรี ν และกำหนดให้ Z และ Y เป็นตัวแปรสุ่มอิสระกันแล้ว ตัวแปรสุ่ม $T = \frac{Z}{\sqrt{Y/\nu}}$ จะมีการแจกแจงแบบที ด้วยระดับขั้นความเสรีเท่ากับ ν ซึ่งมีฟังก์ชันความหนาแน่นความน่าจะเป็นของ T คือ

$$f(t,\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu}\,\Gamma\left(\frac{\nu}{2}\right)\left(1+\frac{t^2}{\nu}\right)^{\frac{\nu+1}{2}}} \qquad ,-\infty < t < \infty,\ \nu > 0$$

โดยที่ ν เป็นพารามิเตอร์แสดงรูปร่างของการแจกแจง (Shape parameter)

ลักษณะทั่วไปของตัวแปรสุ่ม T ที่มีการแจกแจงแบบที

1. ค่าเฉลี่ยของตัวแปรสุ่ม X

$$E(T) = \mu = 0 \qquad , \nu > 1$$

2. ความแปรปรวนของตัวแปรสุ่ม X

$$Var(T) = \frac{\nu}{\nu-2} \qquad , \nu > 2$$

3. สัมประสิทธิ์ความเบ้

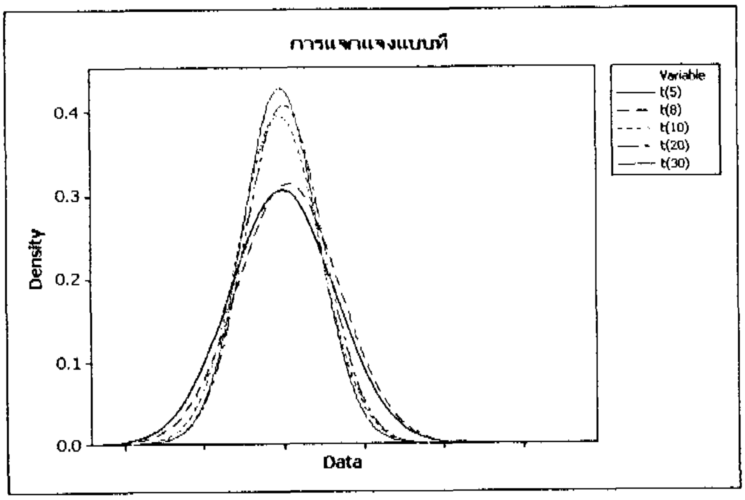
$$\gamma_1 = 0$$

4. สัมประสิทธิ์ความโด่ง

$$\gamma_2 = \frac{3(\nu-2)}{(\nu-4)} \qquad , \nu > 4$$

รูปที่ 2.4

แสดงฟังก์ชันความหนาแน่นของการแจกแจงแบบที



2.2.4 การแจกแจงแบบล็อกนอร์มัล (Lognormal Distribution)

กำหนดให้ X เป็นตัวแปรสุ่มต่อเนื่องที่มีการแจกแจงแบบล็อกนอร์มัล ด้วยพารามิเตอร์ μ และ σ^2 ดังนั้นจะมีฟังก์ชันความหนาแน่นความน่าจะเป็น คือ

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

ลักษณะทั่วไปของตัวแปรสุ่ม X ที่มีการแจกแจงแบบล็อกนอร์มัล

1. ค่าเฉลี่ยของตัวแปรสุ่ม X

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

2. ความแปรปรวนของตัวแปรสุ่ม X

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

โดยที่ $m = \exp(\mu)$

$$\omega = \exp(\sigma^2)$$

3. สัมประสิทธิ์ความเบ้

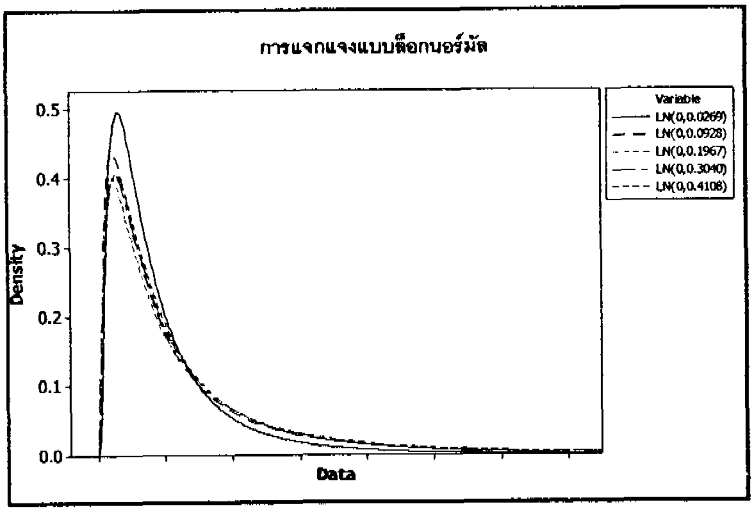
$$\gamma_1 = (\omega + 2)\sqrt{\omega - 1}$$

4. สัมประสิทธิ์ความโด่ง

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

รูปที่ 2.5

แสดงฟังก์ชันความหนาแน่นของการแจกแจงแบบล็อกนอร์มัล



2.2.5 การแจกแจงแบบบีตา (Beta Distribution)

กำหนดให้ X เป็นตัวแปรสุ่มต่อเนื่องที่มีการแจกแจงแบบบีตา ด้วยพารามิเตอร์ α และ β จะมีฟังก์ชันความหนาแน่นความน่าจะเป็นของ X คือ

$$f(x;\alpha,\beta)=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1} \quad , 0<x<1 \ , \alpha>0, \beta>0$$

โดยที่ α และ β เป็นพารามิเตอร์แสดงรูปร่างของการแจกแจง (Shape parameter)

ลักษณะทั่วไปของตัวแปรสุ่ม X ที่มีการแจกแจงแบบบีตา

1. ค่าเฉลี่ยของตัวแปรสุ่ม X

$$E(X)=\frac{\alpha}{\alpha+\beta}$$

2. ความแปรปรวนของตัวแปรสุ่ม X

$$Var(X)=\frac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)^2}$$

3. สัมประสิทธิ์ความเบ้

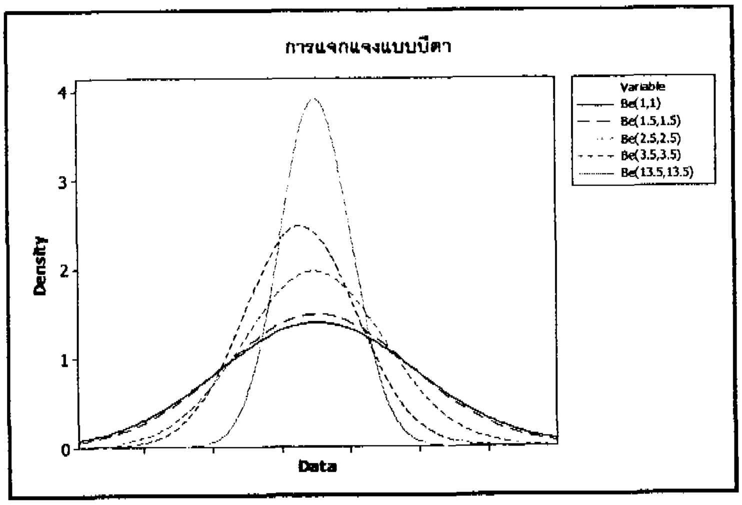
$$\gamma_1=\frac{2(\beta-\alpha)\sqrt{\frac{1}{\alpha}+\frac{1}{\beta}+\frac{1}{\alpha\beta}}}{(\alpha+\beta+2)}$$

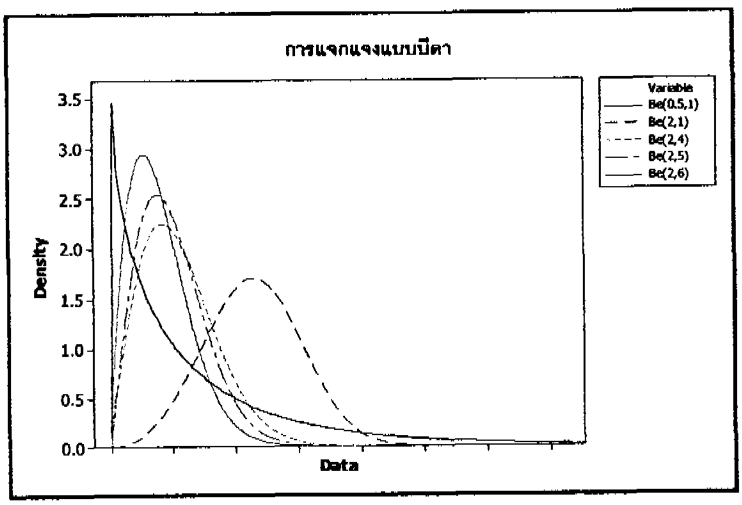
4. สัมประสิทธิ์ความโด่ง

$$\gamma_2=\frac{3(\alpha+\beta+1)[2(\alpha+\beta)^2+\alpha\beta(\alpha+\beta-6)]}{\alpha\beta(\alpha+\beta+2)(\alpha+\beta+3)}$$

รูปที่ 2.6

แสดงฟังก์ชันความหนาแน่นของการแจกแจงแบบบีตา





2.2.6 การแจกแจงแบบไวบูลล์ (Weibull Distribution)

กำหนดให้ X เป็นตัวแปรสุ่มต่อเนื่องที่มีการแจกแจงแบบไวบูลล์ ด้วยพารามิเตอร์ α และ β จะมีฟังก์ชันความหนาแน่นความน่าจะเป็นของ X คือ

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

โดยที่ α เป็นพารามิเตอร์แสดงรูปร่างของการแจกแจง (Shape parameter)

β เป็นพารามิเตอร์แสดงขนาดของการแจกแจง (Scale parameter)

ลักษณะทั่วไปของตัวแปรสุ่ม X ที่มีการแจกแจงแบบไวบูลล์

1. ค่าเฉลี่ยของตัวแปรสุ่ม X

$$E(X) = \beta\Gamma\left(1+\frac{1}{\alpha}\right)$$

2. ความแปรปรวนของตัวแปรสุ่ม X

$$Var(X) = \beta^2\left[\Gamma\left(1+\frac{2}{\alpha}\right)-\Gamma^2\left(1+\frac{1}{\alpha}\right)\right]$$

3. สัมประสิทธิ์ความเบ้

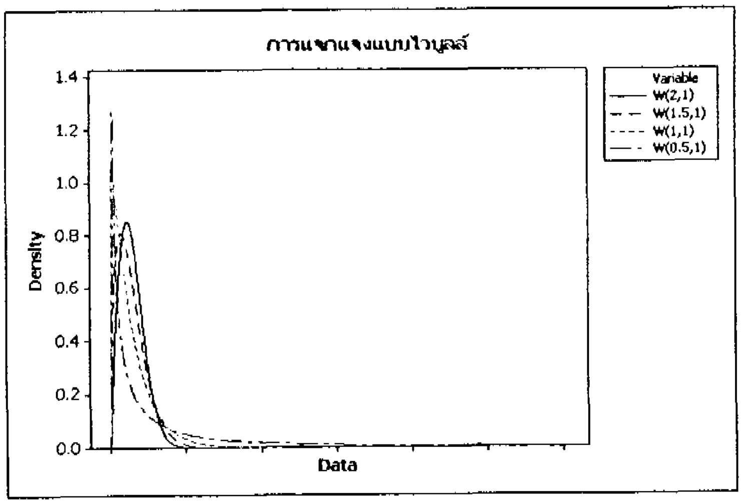
$$\gamma_1 = \frac{\Gamma\left(1+\frac{3}{\alpha}\right)-3\Gamma\left(1+\frac{2}{\alpha}\right)\Gamma\left(1+\frac{1}{\alpha}\right)+2\Gamma^3\left(1+\frac{1}{\alpha}\right)}{\left[\Gamma\left(1+\frac{2}{\alpha}\right)-\Gamma^2\left(1+\frac{1}{\alpha}\right)\right]^{\frac{3}{2}}}$$

4. สัมประสิทธิ์ความโด่ง

$$\gamma_2 = \frac{\Gamma\left(1+\frac{4}{\alpha}\right)+6\Gamma\left(1+\frac{2}{\alpha}\right)\Gamma^2\left(1+\frac{1}{\alpha}\right)-3\Gamma^4\left(1+\frac{1}{\alpha}\right)-4\Gamma\left(1+\frac{3}{\alpha}\right)\Gamma\left(1+\frac{1}{\alpha}\right)}{\left[\Gamma\left(1+\frac{2}{\alpha}\right)-\Gamma^2\left(1+\frac{1}{\alpha}\right)\right]^{2}}$$

รูปที่ 2.7

แสดงฟังก์ชันความหนาแน่นของการแจกแจงแบบไวบูลล์



2.3 ตัวสถิติทดสอบสำหรับการทดสอบการแจกแจงแบบปกติ

กำหนดให้ X เป็นตัวแปรสุ่มต่อเนื่องจากฟังก์ชันการแจกแจง F(x) และ $X_1, X_2, \ldots, X_n$ เป็นตัวอย่างสุ่มขนาด n ในการทดสอบการแจกแจงของประชากรว่าจะมีการแจกแจงแบบใดนั้น จะต้องกำหนดสมมติฐานในการทดสอบ โดยในการวิจัยนี้จะศึกษาการแจกแจงแบบปกติ ดังนั้น สมมติฐานในการทดสอบภาวะสารูปสนิทดีสำหรับการทดสอบการแจกแจงแบบปกติ คือ

สมมติฐานว่าง H_0 : ประชากรมีการแจกแจงแบบปกติ

แย้งกับ H_1 : ประชากรไม่มีการแจกแจงแบบปกติ

2.3.1 ตัวสถิติทดสอบ Anderson-Darling (A^2)

ในปี ค.ศ. 1952 แอนเดอร์สันและดาร์ลิง (Anderson and Darling) ได้เสนอตัวสถิติทดสอบ Anderson-Darling (A^2) ในการทดสอบการแจกแจงแบบปกติ โดยกำหนดให้ X เป็นตัวแปรสุ่มจากฟังก์ชันการแจกแจง F(x) และ $X_1, X_2, \ldots, X_n$ เป็นตัวอย่างสุ่มขนาด n ที่มีสถิติอันดับ $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ ในการทดสอบสมมติฐาน

H_0 : ประชากรมีการแจกแจงแบบปกติ แย้งกับ

H_1 : ประชากรไม่มีการแจกแจงแบบปกติ

โดยมีตัวสถิติทดสอบ คือ

$$A^2 = -n - \frac{1}{n}\left[\sum_{i=1}^{n}(2i-1)\left[\ln F(X_{(i)}) + \ln\left(1-F(X_{(n-i+1)})\right)\right]\right]$$

โดยที่ $F(X_{(i)})$ คือ ความน่าจะเป็นสะสมของการแจกแจงแบบปกติมาตรฐาน

n คือ ขนาดตัวอย่าง

เกณฑ์การตัดสินใจในการทดสอบสมมติฐานที่ระดับนัยสำคัญ α คือจะปฏิเสธสมมติฐานว่าง H_0 เมื่อค่า A^2 ที่คำนวณได้มีค่ามากกว่าค่าวิกฤตที่ได้จากตาราง ก1 Percentage points of Anderson-Darling test (ภาคผนวก ก) ณ ขนาดตัวอย่าง และระดับนัยสำคัญที่กำหนด

2.3.2 ตัวสถิติทดสอบ Shapiro-Wilk (W)

ในปี ค.ศ. 1965 ชาร์พิโรและวิลค์ (Shapiro and Wilk) ได้เสนอตัวสถิติทดสอบ Shapiro-Wilk (W) ในการทดสอบการแจกแจงแบบปกติ โดยกำหนดให้ X เป็นตัวแปรสุ่มจากฟังก์ชันการแจกแจง F(x) และ $X_1, X_2, \ldots, X_n$ เป็นตัวอย่างสุ่มขนาด n ที่มีสถิติอันดับ $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ ในการทดสอบสมมติฐาน

H_0 : ประชากรมีการแจกแจงแบบปกติ แย้งกับ

H_1 : ประชากรไม่มีการแจกแจงแบบปกติ

โดยมีตัวสถิติทดสอบ คือ

$$W = \frac{\left[\sum_{i=1}^{k} a_{n-i+1}\left(x_{(n-i+1)} - x_{(i)}\right)\right]^2}{\sum_{i=1}^{n}\left(x_{(i)} - \bar{x}\right)^2}$$

โดยที่ a_{n-i+1} คือ ค่าสัมประสิทธิ์ที่ได้จากตาราง ก2 (ภาคผนวก ก)

k คือ จำนวนเต็มเล็กที่สุดที่มากกว่าหรือเท่ากับ n/2

n คือ ขนาดตัวอย่าง

เกณฑ์การตัดสินใจในการทดสอบสมมติฐานที่ระดับนัยสำคัญ α คือจะปฏิเสธสมมติฐานว่าง H_0 เมื่อค่า W ที่คำนวณได้มีค่าน้อยกว่าค่าวิกฤตที่ได้จากตาราง ก3 Percentage points of Shapiro-Wilk test (ภาคผนวก ก) ณ ขนาดตัวอย่าง และระดับนัยสำคัญที่กำหนด

2.3.3 ตัวสถิติทดสอบ Shapiro-Francia (W')

ในปี ค.ศ. 1972 ชาร์พิโรและฟรานเซีย (Shapiro and Francia) ได้เสนอตัวสถิติทดสอบ Shapiro-Francia (W') ในการทดสอบการแจกแจงแบบปกติ โดยกำหนดให้ X เป็นตัวแปรสุ่มจากฟังก์ชันการแจกแจง F(x) และ $X_1, X_2, \ldots, X_n$ เป็นตัวอย่างสุ่มขนาด n ที่มีสถิติอันดับ $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ ในการทดสอบสมมติฐาน

H_0 : ประชากรมีการแจกแจงแบบปกติ แย้งกับ

H_1 : ประชากรไม่มีการแจกแจงแบบปกติ

โดยมีตัวสถิติทดสอบ คือ

$$W' = \frac{\left[\sum_{i=1}^{n} m_i X_{(i)}\right]^2}{\sum_{i=1}^{n} m_i^2 \cdot \sum_{i=1}^{n} (X_{(i)} - \bar{X})^2}$$

โดยที่ m_i คือ ค่าคาดหวังของสถิติอันดับจากการแจกแจงแบบปกติมาตรฐาน (Expected value of normal order statistics) ได้จากตาราง ก4 (ภาคผนวก ก)

n คือ ขนาดตัวอย่าง

เกณฑ์การตัดสินใจในการทดสอบสมมติฐานที่ระดับนัยสำคัญ α คือจะปฏิเสธสมมติฐานว่าง H_0 เมื่อค่า W' ที่คำนวณได้มีค่าน้อยกว่าค่าวิกฤตที่ได้จากตาราง ก5 Percentage points of Shapiro-Francia test (ภาคผนวก ก) ณ ขนาดตัวอย่าง และระดับนัยสำคัญที่กำหนด

2.3.4 ตัวสถิติทดสอบของชางและวู

ในปี ค.ศ. 2005 ชางและวู (Zhang and Wu) ได้ทำการศึกษาตัวสถิติทดสอบที่ได้จากอัตราส่วนภาวะน่าจะเป็น (Likelihood ratio test statistic) คือ Z_A , Z_C และ Z_K เพิ่มเติมจากการศึกษาของ ชาง (Zhang, 2002) เพื่อใช้สำหรับการทดสอบการแจกแจงแบบปกติ โดยกำหนดให้ X เป็นตัวแปรสุ่มจากฟังก์ชันการแจกแจง F(x) และ $X_1, X_2, \ldots, X_n$ เป็นตัวอย่างสุ่มขนาด n ที่มีสถิติอันดับ $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ ในการทดสอบสมมติฐาน

H_0 : ประชากรมีการแจกแจงแบบปกติ แย้งกับ

H_1 : ประชากรไม่มีการแจกแจงแบบปกติ

โดยชางและวูได้เสนอตัวสถิติทดสอบ 3 ตัว คือ

$$Z_A = -\sum_{i=1}^{n}\left[\frac{\log F_0(X_{(i)})}{n-i+0.5}+\frac{\log\left[1-F_0(X_{(i)})\right]}{i-0.5}\right]$$

$$Z_c = \sum_{i=1}^{n}\left[\log\left(\frac{\frac{F_0(X_{(i)})^{-1}-1}{(n-0.5)}}{\frac{}{(i-0.75)}-1}\right)\right]^2$$

$$Z_K = \max_{1\le i\le n}\left[(i-0.5)\log\left[\frac{i-0.5}{nF_0(X_{(i)})}\right]+[n-i+0.5]\log\frac{n-i+0.5}{n\left(1-F_0(X_{(i)})\right)}\right]$$

โดยที่ $F_0(X_{(i)})$ คือ ความน่าจะเป็นสะสมของการแจกแจงแบบปกติมาตรฐาน

n คือ ขนาดตัวอย่าง

เกณฑ์การตัดสินใจในการทดสอบสมมติฐานที่ระดับนัยสำคัญ α ของตัวสถิติ Z_A คือจะปฏิเสธสมมติฐานว่าง H_0 เมื่อค่า $10Z_A$ - 32 ที่คำนวณได้มีค่ามากกว่าค่าวิกฤตที่ได้จากตาราง ก6 Percentage points of $10Z_A - 32$ for testing normality (ภาคผนวก ก) ณ ขนาดตัวอย่างและระดับนัยสำคัญที่กำหนด

เกณฑ์การตัดสินใจในการทดสอบสมมติฐานที่ระดับนัยสำคัญ α ของตัวสถิติทดสอบ Z_C คือจะปฏิเสธสมมติฐานว่าง H_0 เมื่อค่า Z_C ที่คำนวณได้มีค่ามากกว่าค่าวิกฤตที่ได้จาก

ตาราง ก7 Percentage points of Z_C for testing normality (ภาคผนวก ก) ณ ขนาดตัวอย่าง และระดับนัยสำคัญที่กำหนด

เกณฑ์การตัดสินใจในการทดสอบสมมติฐานที่ระดับนัยสำคัญ α ของตัวสถิติทดสอบ Z_K คือจะปฏิเสธสมมติฐานว่าง H_0 เมื่อค่า Z_K ที่คำนวณได้มีค่ามากกว่าค่าวิกฤตที่ได้จากตาราง ก8 Percentage points of Z_K for testing normality (ภาคผนวก ก) ณ ขนาดตัวอย่าง และระดับนัยสำคัญที่กำหนด

<u>การพิสูจน์การหาตัวสถิติทดสอบของซาง</u>

กำหนดให้ X เป็นตัวแปรสุ่มต่อเนื่องจากฟังก์ชันการแจกแจง F(x) และ $X_1, X_2, \ldots, X_n$ เป็นตัวอย่างสุ่มขนาด n ที่มีสถิติอันดับ $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ ในการทดสอบสมมติฐาน

$H_0: F(x) = F_0(x)$,สำหรับทุกค่า $x \in (-\infty, \infty)$ แย้งกับ

$H_1: F(x) \neq F_0(x)$,สำหรับบางค่า $x \in (-\infty, \infty)$

เมื่อ $F_0(x)$ เป็นฟังก์ชันการแจกแจงที่กำหนดในการทดสอบ

โดยใช้ t เป็นจุดแบ่ง x ให้มีการแจกแจงแบบเบอร์นูลลี (Bernoulli distribution) และทำการทดสอบ t ครั้ง แล้วทดสอบผลร่วมที่ได้ทั้งหมดที่เป็นไปได้โดยมีสมมติฐานที่ทดสอบดังนี้

ให้
$$H_0 = \bigcap_{t \in (-\infty, \infty)} H_{0t}$$
$$H_1 = \bigcup_{t \in (-\infty, \infty)} H_{1t}$$

เมื่อ $H_{0t}: F(t) = F_0(t)$ และ $H_{1t}: F(t) \neq F_0(t)$

ดังนั้นการทดสอบสมมติฐาน H_0 แย้งกับ H_1 จะสมมูลกับ H_{0t} แย้งกับ H_{1t} สำหรับทุก $t \in (-\infty, \infty)$

การทดสอบสมมติฐาน H_{0t} แย้งกับ H_{1t} เมื่อ t คงที่ จะทำให้ตัวอย่างสุ่มแบ่งเป็น 2 ส่วนภายใต้ฟังก์ชันอินดิเคเตอร์ (Indicator Function) $x_{it} = I(x_i \leq t)$, $(i = 1,2,\ldots,n)$ ดังนั้น $P(x_{it} = 1) = F(t)$ และ $P(x_{it} = 0) = 1 - F(t)$ เพราะฉะนั้น $x_{it} \sim Ber(F(t))$

ให้ Z_t เป็นตัวสถิติทดสอบที่ใช้ทดสอบสมมติฐาน H_{0t} แย้งกับ H_{0t} กำหนดตัวสถิติทดสอบ[2] Z ในการทดสอบ $H_0 = \bigcap_{t\in(-\infty,\infty)} H_{0t}$ และ $H_1 = \bigcup_{t\in(-\infty,\infty)} H_{1t}$ ดังนี้

$$Z = \int_{-\infty}^{\infty} Z_t \, dw(t) \quad \text{และ} \tag{2.1}$$

$$Z_{max} = \sup_{t\in(-\infty,\infty)} \left[Z_t w(t)\right] \tag{2.2}$$

เมื่อ w(t) คือฟังก์ชันถ่วงน้ำหนัก (weight function) และจะปฏิเสธสมมติฐาน ถ้า Z หรือ Z_{max} มีค่ามาก

ดังนั้นสามารถหาตัวสถิติทดสอบอัตราส่วนภาวะน่าจะเป็น (Likelihood ratio test statistic) หรือ G_t^2 ในการทดสอบสมมติฐาน H_{0t} แย้งกับ H_{1t} ได้ดังนี้

จาก $x \sim Ber(F(t))$

ดังนั้น $f(x) = F(t)^x [1-F(t)]^{1-x} \quad , x = 0,1$

$$\Omega = \{F(t) \mid ,0 \le F(t) \le 1\}$$

$$\omega = \{F_0(t) \mid ,0 \le F_0(t) \le 1\}$$

$$L(\hat{\Omega}) = F_n(t)^{\sum_{i=1}^{n} x_i} \left[1 - F_n(t)\right]^{n-\sum_{i=1}^{n} x_i}$$

$$L(\hat{\omega}) = F_0(t)^{\sum_{i=1}^{n} x_i} \left[1 - F_0(t)\right]^{n-\sum_{i=1}^{n} x_i}$$

ดังนั้นจะได้ G_t^2

$$G_t^2 = -2\log\left[\frac{L(\hat{\omega})}{L(\hat{\Omega})}\right]$$

$$= 2\log\left[\frac{L(\hat{\Omega})}{L(\hat{\omega})}\right]$$

$$= 2\log\left(\frac{F_n(t)^{\sum_{i=1}^{n} x_i} \left[1 - F_n(t)\right]^{n-\sum_{i=1}^{n} x_i}}{F_0(t)^{\sum_{i=1}^{n} x_i} \left[1 - F_0(t)\right]^{n-\sum_{i=1}^{n} x_i}}\right)$$

---

[2] Jin Zhang, 'Powerful goodness-of-fit tests based on the likelihood ratio', Journal of the Royal Statistical Society Series B 64(2002) : p.282

$$= \quad 2\log\left[\left(\frac{F_n(t)}{F_o(t)}\right)^{nF_n(t)}\left(\frac{1-F_n(t)}{1-F_o(t)}\right)^{n-nF_n(t)}\right]$$

$$= \quad 2\left[nF_n(t)\log\left(\frac{F_n(t)}{F_o(t)}\right) + n\left(1-F_n(t)\right)\log\left(\frac{1-F_n(t)}{1-F_o(t)}\right)\right]$$

$$G_t^2 \quad = \quad 2n\left[F_n(t)\log\left(\frac{F_n(t)}{F_o(t)}\right) + \left(1-F_n(t)\right)\log\left(\frac{1-F_n(t)}{1-F_o(t)}\right)\right]$$

เมื่อ $F_n(t)$ เป็นฟังก์ชันการแจกแจงของตัวอย่าง (ค่าสังเกต) (Empirical distribution function)

ดังนั้นจะสร้างตัวสถิติทดสอบที่มีการแจกแจงสมมาตร จากการเลือกฟังก์ชันถ่วงน้ำหนักที่เหมาะสมให้กับอัตราส่วนภาวะน่าจะเป็น และนอกจากการเลือกฟังก์ชันแล้วบางครั้งจำเป็นจะต้องมีการปรับ $F_n(t)$ ที่จุด $X_{(i)}$ ให้เป็นฟังก์ชันต่อเนื่อง โดยนิยามเป็น $F_n(X_{(i)}) = \frac{i-c}{n+1-2c}$ เมื่อ c เป็นค่าคงที่อยู่ระหว่าง 0 และ 1 แต่โดยทั่วไปแล้วจะเลือก c = ½

$$\text{ดังนั้นจะได้ } F_n(X_{(i)}) = \frac{i-\frac{1}{2}}{n+1-2\left(\frac{1}{2}\right)} = \frac{i-0.5}{n}$$

1. เลือก $dw(t) = F_n(t)^{-1}\left[1-F_n(t)\right]^{-1}dF_n(t)$

แทนค่า Z_t ในสมการ (2.1) ด้วย G_t^2 จะได้

$$Z \quad = \quad \int_{-\infty}^{\infty} G_t^2\, dw(t)$$

$$= \quad \int_{-\infty}^{\infty} 2n\left[F_n(t)\log\frac{F_n(t)}{F_o(t)} + \left[1-F_n(t)\right]\log\frac{1-F_n(t)}{1-F_o(t)}\right].F_n(t)^{-1}\left[1-F_n(t)\right]^{-1}dF_n(t)$$

$$= \quad 2n\int_{-\infty}^{\infty}\left[\frac{1}{1-F_n(t)}\log\frac{F_n(t)}{F_o(t)} + \frac{1}{F_n(t)}\log\frac{1-F_n(t)}{1-F_o(t)}\right]dF_n(t)$$

$$= \frac{1}{n}\left[2n\sum_{i=1}^{n}\left[\frac{1}{1-\frac{i-0.5}{n}}\log\left[\frac{\frac{i-0.5}{n}}{F_0(X_{(i)})}\right]+\frac{1}{\frac{i-0.5}{n}}\log\left[\frac{1-\frac{i-0.5}{n}}{1-F_0(X_{(i)})}\right]\right]\right]$$

$$= 2\sum_{i=1}^{n}\left[\frac{n}{n-i+0.5}\log\frac{i-0.5}{nF_0(X_{(i)})}+\frac{n}{i-0.5}\log\frac{n-i+0.5}{n\left(1-F_0(X_{(i)})\right)}\right]$$

ซึ่งสมมูลกับ

$$Z = -\sum_{i=1}^{n}\left[\frac{\log F_0(X_{(i)})}{n-i+0.5}+\frac{\log\left[1-F_0(X_{(i)})\right]}{i-0.5}\right]$$

จะเรียกตัวสถิติทดสอบที่ได้นี้ว่า Z_A

2. เลือก $dw(t) = F_0(t)^{-1}\left[1-F_0(t)\right]^{-1}dF_0(t)$

แทนค่า Z_t ในสมการ (2.1) ด้วย G_t^2 จะได้

$$Z = \int_{-\infty}^{\infty} G_t^2\,dw(t)$$

$$= \int_{-\infty}^{\infty} 2n\left[F_n(t)\log\frac{F_n(t)}{F_0(t)}+\left[1-F_n(t)\right]\log\frac{1-F_n(t)}{1-F_0(t)}\right].F_0(t)^{-1}\left[1-F_0(t)\right]^{-1}dF_0(t)$$

$$= 2n\int_{-\infty}^{\infty}\left[\frac{F_n(t)}{F_0(t)\left(1-F_0(t)\right)}.\log\frac{F_n(t)}{F_0(t)} + \frac{1-F_n(t)}{F_0(t)\left(1-F_0(t)\right)}.\log\frac{1-F_n(t)}{1-F_0(t)}\right]dF_0(t)$$

$$= \frac{1}{n}\left[2n\sum_{i=1}^{n}\left[\frac{\frac{i-0.5}{n}}{F_0(X_{(i)})\left(1-F_0(X_{(i)})\right)}\log\left(\frac{\frac{i-0.5}{n}}{F_0(X_{(i)})}\right)+\frac{1-\frac{i-0.5}{n}}{F_0(X_{(i)})\left(1-F_0(X_{(i)})\right)}\log\left(\frac{1-\frac{i-0.5}{n}}{1-F_0(X_{(i)})}\right)\right]\right]$$

ซึ่งสมมูลกับ $\sum_{i=1}^{n}\left[\log\left(F_0(X_{(i)})^{-1}-1\right)-b_{i-1}+b_i\right]^2 + C_n$

เมื่อ C_n เป็นค่าคงที่ และ $b_i = i\log\left(\frac{1}{n}\right) + (n-i)\log\left(1-\frac{i}{n}\right)$

ซึ่ง $b_{i-1} - b_i \approx \log\left(\frac{n-0.5}{i-0.75} - 1\right)$

ดังนั้น

$$Z \approx \sum_{i=1}^{n}\left[\log\left(F_0(X_{(i)})^{-1} - 1\right) - \log\left(\frac{n-0.5}{i-0.75} - 1\right)\right]^2 + C_n$$

$$\approx \sum_{i=1}^{n}\left[\log\left(\frac{F_0(X_{(i)})^{-1} - 1}{\frac{(n-0.5)}{(i-0.75)} - 1}\right)\right]^2$$

$$= Z_c$$

3. เลือก w(t) = 1

ให้ $X_{(0)} = -\infty$ และ $X_{(n+1)} = \infty$ แทนค่า Z_t ในสมการ (2.2) ด้วย G_t^2 จะได้

$$Z_{max} = \sup_{t\in(-\infty,\infty)} 2n\left[F_n(t)\log\left[\frac{F_n(t)}{F_0(t)}\right] + [1-F_n(t)]\log\frac{1-F_n(t)}{1-F_0(t)}\right]$$

$$= \sup_{t\in(-\infty,\infty)} 2n\left[\frac{i-0.5}{n}\log\left[\frac{F_n(t)}{F_0(t)}\right] + \left[1-\frac{i-0.5}{n}\right]\log\frac{1-F_n(t)}{1-F_0(t)}\right]$$

$$= \max_{1\le i\le n}\left[(i-0.5)\log\left[\frac{i-0.5}{nF_0(X_{(i)})}\right] + [n-i+0.5]\log\frac{n-i+0.5}{n\left(1-F_0(X_{(i)})\right)}\right]$$

จะเรียกตัวสถิติทดสอบที่ได้นี้ว่า Z_K

2.4 ผลงานวิจัยที่เกี่ยวข้อง

ชาพิโรและวิลค์ (Shapiro and Wilk, 1965) ได้เสนอตัวสถิติทดสอบ Shapiro-Wilk (W) ซึ่งเป็นตัวสถิติทดสอบในการทดสอบสารูปสนิทดีสำหรับการแจกแจงแบบปกติ เมื่อขนาดตัวอย่างน้อยกว่าหรือเท่ากับ 50 โดยนำค่าสัมประสิทธิ์เชิงเส้นที่เหมาะสมกับข้อมูลที่มีการเรียงลำดับ (Order Sample) มาใช้ในการคำนวณ

ชาพิโรและคณะ (Shapiro, Wilk and Chen, 1968) เป็นผู้ริเริ่มในการศึกษากำลังการทดสอบของตัวสถิติทดสอบต่างๆ สำหรับการทดสอบการแจกแจงแบบปกติ โดยทำการศึกษาตัวสถิติทดสอบ 9 ตัว คือ ตัวสถิติทดสอบ Shapiro-Wilk (W) ตัวสถิติทดสอบ Standard Third Moment ($\sqrt{b_1}$) ตัวสถิติทดสอบ Standard Fourth Moment (b_2) ตัวสถิติทดสอบ Kolmogorov-Smirnov (K) ตัวสถิติทดสอบ Cramer-von Mises (W^2) ตัวสถิติทดสอบ Anderson-Darling (A^2) ตัวสถิติทดสอบ Durbin ตัวสถิติทดสอบ Chi-square (χ^2) และตัวสถิติทดสอบ Studentized Range (u) ได้ผลสรุปโดยย่อดังนี้ตัวสถิติทดสอบ W ใช้ได้ดีในการทดสอบทั่วไป ตัวสถิติทดสอบ Studentized Range (u) มีกำลังการทดสอบสูงเมื่อประชากรมีการแจกแจงแบบสมมาตรหางสั้น (Symmetric Short-Tailed) และมีกำลังการทดสอบต่ำเมื่อประชากรมีการแจกแจงแบบเบ้

ชาพิโรและฟรานเชีย (Shapiro and Francia, 1972) ได้เสนอตัวสถิติทดสอบ Shapiro-Francia (W') สำหรับการทดสอบการแจกแจงแบบปกติ โดยใช้ผลต่างกำลังสองระหว่างสัมประสิทธิ์สหสัมพันธ์เพียร์สันของข้อมูลที่มีการเรียงลำดับ (Order Sample) กับค่าคาดหวังของการแจกแจงปกติมาตรฐานที่มีการเรียงลำดับ (Normal score)

ฟิลลิเบน (Filliben, 1975) ได้เสนอตัวสถิติทดสอบ Probability Plot Correlation Coefficient หรือตัวสถิติทดสอบ Filliben (r) ซึ่งเป็นตัวสถิติที่ใช้ในการทดสอบสารูปสนิทดีสำหรับการแจกแจงแบบปกติ โดยดัดแปลงมาจากตัวสถิติทดสอบ Shapiro-Wilk (W) และได้ทำการเปรียบเทียบกำลังการทดสอบของตัวสถิติทดสอบที่ใช้ในการทดสอบการแจกแจงแบบปกติ 8 ตัว คือ ตัวสถิติทดสอบ Standard Third Moment ($\sqrt{b_1}$) ตัวสถิติทดสอบ Standard Fourth Moment (b_2) ตัวสถิติทดสอบ a ตัวสถิติทดสอบ Studentized Range (u) ตัวสถิติทดสอบ Shapiro-Wilk (W) ตัวสถิติทดสอบ Shapiro-Francia (W') ตัวสถิติทดสอบ D'Agostino (D) และตัวสถิติทดสอบ Filliben (r) พบว่า ตัวสถิติทดสอบ Filliben (r) มีกำลังการทดสอบสูงกว่าตัวสถิติทดสอบ

Shapiro-Wilk (W) เมื่อประชากรมีการแจกแจงแบบใกล้เคียงปกติ (Near Normal) และสมมาตรหางยาว (Symmetric Long-Tailed)

เวอริลและจอห์นสัน (Verrill and Johnson, 1980) ได้ศึกษาเปรียบเทียบกำลังการทดสอบสำหรับการทดสอบการแจกแจงแบบปกติในกรณีข้อมูลถูกตัดทิ้ง ของตัวสถิติทดสอบ Cramer-Von Mises (W^2), ตัวสถิติทดสอบ Anderson-Darling (A^2), ตัวสถิติทดสอบ Shapiro-Wilk (W), ตัวสถิติทดสอบ Shapiro-Francia (W'), ตัวสถิติทดสอบ Dewet-Venter และตัวสถิติทดสอบ Chi-square พบว่าตัวสถิติทดสอบ Shapiro-Wilk (W) มีกำลังการทดสอบสูงสุดในกรณีข้อมูลสมบูรณ์ มีลักษณะสมมาตร และหางยาว ส่วนกำลังการทดสอบของตัวสถิติทดสอบ Cramer-Von Mises (W^2), ตัวสถิติทดสอบ Anderson-Darling (A^2) และตัวสถิติทดสอบ Shapiro-Francia (W') มีค่าใกล้เคียงกัน

สมพิศ โชติวิทยธารากร (2530) ได้ศึกษาเปรียบเทียบกำลังการทดสอบของตัวสถิติบางตัวที่ใช้ในการทดสอบการแจกแจงปกติ ได้แก่ ตัวสถิติไคกำลังสอง (χ^2), ตัวสถิติ Studentized Range (u), ตัวสถิติ Shapiro-Wilk, ตัวสถิติ Filliben และตัวสถิติ Hannu Oja (T'_1 และ T'_2) ภายใต้การแจกแจงแบบปกติ การแจกแจงปกติปลอมปน และแบบเบ้ ขนาดตัวอย่างเท่ากับ 10, 30, 50 และ 100 ตามลำดับ และกำหนดระดับนัยสำคัญของการทดสอบเท่ากับ 0.01, 0.05 และ 0.10 ซึ่งพบว่าในการเลือกใช้ตัวสถิติสำหรับการทดสอบการแจกแจงปกติ ควรเลือกใช้ตัวสถิติ Shapiro-Wilk เมื่อขนาดตัวอย่างน้อยกว่าหรือเท่ากับ 50 เนื่องจากให้กำลังการทดสอบสูงในสถานการณ์ต่างๆ และสามารถควบคุมความน่าจะเป็นที่จะเกิดความผิดพลาดประเภทที่ 1 ได้ดี แต่เมื่อขนาดตัวอย่างมากกว่า 50 และประชากรมีการแจกแจงแบบหางยาวซึ่งมีความเบ้น้อยกว่าหรือเท่ากับ 0.50 และความโด่งมีค่ามาก ควรเลือกใช้ตัวสถิติ Filliben

เกตุจันทร์ พัชรินทร์ศักดิ์ (2534) ได้ศึกษาเปรียบเทียบวิธีการนอนพาราเมตริกสำหรับการทดสอบการแจกแจงแบบปกติ ได้แก่ ตัวสถิติ Shapiro-Wilk, ตัวสถิติ Cramer-von Mises, ตัวสถิติ Anderson-Darling, ตัวสถิติ Watson, ตัวสถิติ Kuiper และตัวสถิติ Dubin ภายใต้การแจกแจงแบบปกติ และการแจกแจงแบบเบ้ ขนาดตัวอย่างเท่ากับ 10, 20, 30, 50 และ 100 ตามลำดับ และกำหนดระดับนัยสำคัญของการทดสอบเท่ากับ 0.01, 0.05 และ 0.10 ซึ่งพบว่าในการเลือกใช้ตัวสถิติสำหรับการทดสอบการแจกแจงปกติควรเลือกใช้ตัวสถิติ Shapiro-Wilk เมื่อขนาดตัวอย่างน้อยกว่าหรือเท่ากับ 50 แต่หากขนาดตัวอย่างมากกว่า 50 ควรใช้สถิติ Anderson-Darling เนื่องจากสามารถควบคุมความน่าจะเป็นที่จะเกิดความผิดพลาดประเภทที่ 1 ได้ดี และมีกำลังการทดสอบสูง และใช้ตัวสถิติ Kuiper สำหรับประชากรที่มีการแจกแจงแบบสมมาตรหางสั้น ขนาด

ตัวอย่างน้อยกว่าหรือเท่ากับ 30 ใช้ตัวสถิติ Cramer-von Mises สำหรับประชากรที่มีการแจกแจงแบบไม่สมมาตร และจะใช้ตัวสถิติ Watson สำหรับประชากรที่มีการแจกแจงแบบสมมาตร

รวมพร เรืองโรจน์ (2543) ได้ศึกษาเปรียบเทียบกำลังการทดสอบของวิธีสถิติทดสอบการแจกแจงปกติระหว่างไคสแควร์ (Chi-square), สถิติชาพิโร-วิลค์ (Shapiro-Wilk) และสถิติชาพิโร-ฟรานเซีย (Shapiro-Francia) ภายใต้การแจกแจงแบบปกติ และการแจกแจงแบบปกติปลอมปน โดยมีสเกลแฟคเตอร์เป็น 3, 5, 7 และ 10 และเปอร์เซ็นการปลอมปนเป็น 5%, 10%, 20% และ 30% ขนาดตัวอย่างเท่ากับ 10, 30, 50 และ 100 ตามลำดับ และกำหนดระดับนัยสำคัญของการทดสอบเท่ากับ 0.01, 0.05 และ 0.10 พบว่าไม่สามารถทดสอบการแจกแจงแบบปกติได้หากขนาดตัวอย่างมีค่าน้อยกว่า 10 และเสนอว่าในการใช้ตัวสถิติสำหรับการทดสอบการแจกแจงแบบปกติ เมื่อตัวอย่างมีขนาดเท่ากับ 30 ควรเลือกใช้ตัวสถิติ Shapiro-Wilk และเมื่อตัวอย่างมีขนาดมากกว่าเท่ากับ 30 ควรเลือกใช้ตัวสถิติ Shapiro-Francia เนื่องจากสามารถควบคุมความน่าจะเป็นที่จะเกิดความผิดพลาดประเภทที่ 1 ได้ และมีกำลังการทดสอบสูง

ชาง (Zhang, 2002) เสนอตัวสถิติทดสอบ 3 ตัวคือ Z_A , Z_C และ Z_K ในการทดสอบภาวะสารูปสนิทดี (Goodness of fit test) สำหรับการทดสอบแจกแจงที่กำหนด โดยตัวสถิติที่เสนอนั้นอาศัยพื้นฐานจากอัตราส่วนภาวะน่าจะเป็น (Likelihood ratio test statistic) และมีการให้ค่าถ่วงน้ำหนักที่เหมาะสมให้กับ อัตราส่วนภาวะน่าจะเป็นเพื่อสร้างตัวสถิติ และทำการศึกษาเปรียบเทียบตัวสถิติที่ได้กับตัวสถิติอื่นๆ ในการทดสอบภาวะสารูปสนิทดี พบว่าตัวสถิติทดสอบคือ Z_A , Z_C และ Z_K นั้นให้กำลังการทดสอบสูงกว่าตัวสถิติทดสอบ Cramer-von Mises และตัวสถิติทดสอบ Kolmogorov-Smirnov

ชางและวู (Zhang and Wu, 2005) ได้ทำการศึกษาตัวสถิติทดสอบที่ได้จากอัตราส่วนภาวะน่าจะเป็น (Likelihood ratio test statistic) คือ Z_A , Z_C และ Z_K เพิ่มเติมจากการศึกษาของชาง (Zhang, 2002) เพื่อใช้สำหรับการทดสอบการแจกแจงแบบปกติ และเสนอว่าในการใช้สถิติทดสอบทั้ง 3 นี้ จะต้องทำการประมาณค่าพาราเตอร์เสมอถึงแม้ว่าจะทราบค่าพารามิเตอร์ของประชากรหรือไม่ก็ตาม จากการเปรียบเทียบตัวสถิติที่ได้นี้สำหรับการทดสอบการแจกแจงแบบปกติกับตัวสถิติ D'Agotino (D), ตัวสถิติ Anderson-Darling (A^2), ตัวสถิติ Shapiro-Wilk (W) ซึ่งพบว่าตัวสถิติ Z_A ; Z_C และ Z_K นั้นมีกำลังการทดสอบสูงกว่าตัวสถิติ D'Agotino (D)

ชิดชนก ชาญณรงค์ (2548) ได้ศึกษาเปรียบเทียบกำลังการทดสอบของวิธีการทดสอบการแจกแจงแบบปกติ 4 วิธี ได้แก่ ตัวสถิติ Q, ตัวสถิติ D, ตัวสถิติ Kolmogorov และ

Smirnov แบบ two-stage delta-corrected และตัวสถิติ Anderson-Darling ภายใต้การแจกแจงแบบปกติ และการแจกแจงแบบเบ้ ขนาดตัวอย่างเท่ากับ 10, 20, 30, 50, 70, และ 100 ตามลำดับ และกำหนดระดับนัยสำคัญของการทดสอบเท่ากับ 0.01, 0.05 และ 0.10 เสนอว่าในการเลือกใช้ตัวสถิติสำหรับการทดสอบการแจกแจงปกติ เมื่อตัวอย่างมีขนาดมากกว่าเท่ากับ 30 ควรเลือกใช้ตัวสถิติ Anderson-Darling เนื่องจากสามารถควบคุมความน่าจะเป็นที่จะเกิดความผิดพลาดประเภทที่ 1 ได้ดีทุกกรณี และมีกำลังการทดสอบค่อนข้างสูง