บทที่ 2

ทฤษฎีและสถิติที่เกี่ยวข้อง

งานวิจัยมีจุดมุ่งหมายเพื่อหาค่าขนาดตัวอย่างที่จะสุ่มจากประชากรเพื่อประมาณการแจกแจงของตัวสถิติทดสอบที เป็นการแจกแจงที ที่ใช้สำหรับทดสอบสมมติฐานเกี่ยวกับค่าเฉลี่ยประชากร กรณีไม่ทราบค่าเบี่ยงเบนมาตรฐานของประชากร และประชากรมีการแจกแจงที่ไม่ใช่การแจกแจงปกติ โดยผู้วิจัยได้ทำการศึกษาการแจกแจงของประชากรใน 4 ลักษณะคือ การแจกแจงเอกรูป การแจกแจงสมมาตรชนิดหางยาว ได้แก่การแจกแจงโลจิสติค และการแจกแจงที การแจกแจงที่มีความเบ้ ได้แก่ การแจกแจงไคกำลังสอง และการแจกแจงลอกนอร์มัล และการแจกแจงแลมดาของตูกีร์ ซึ่งเป็นการแจกแจงที่กำหนดโดยค่าความเบ้และความโด่งของข้อมูล สำหรับการประมาณการแจกแจงของตัวสถิติทดสอบทีในการการวิจัยครั้งนี้มีหลักเกณฑ์ในการพิจารณา คือ ความสามารถในการควบคุมความผิดพลาดประเภทที่ 1 จากการทดลอง โดยใช้วิธีการเปรียบเทียบค่าระดับนัยสำคัญที่แท้จริงกับค่าระดับนัยสำคัญที่ประมาณการได้ และการทดสอบประสิทธิภาพของผลการวิจัยที่ได้ โดยใช้การทดสอบเทียบความกลมกลืนกัน ซึ่งรายละเอียดและขั้นตอนการประมาณจะกล่าวไว้ในบทที่ 3 สำหรับในบทนี้จะกล่าวรายละเอียดต่างๆ ดังนี้

1. ความเบ้ (skewness) และความโด่ง (kurtosis) เพื่อใช้สำหรับกำหนดรูปแบบการแจกแจงที่กำหนด

2. การแจกแจงต่างๆ ที่กล่าวในงานวิจัยนี้ ซึ่งได้แก่ การแจกแจงเอกรูป การแจกแจงโลจิสติค การแจกแจงที การแจกแจงไคกำลังสอง การแจกแจงลอกนอร์มัล และการแจกแจงแลมดาของตูกีร์ ซึ่งรวมถึงความสัมพันธ์ระหว่างการแจกแจงปกติกับการแจกแจงที

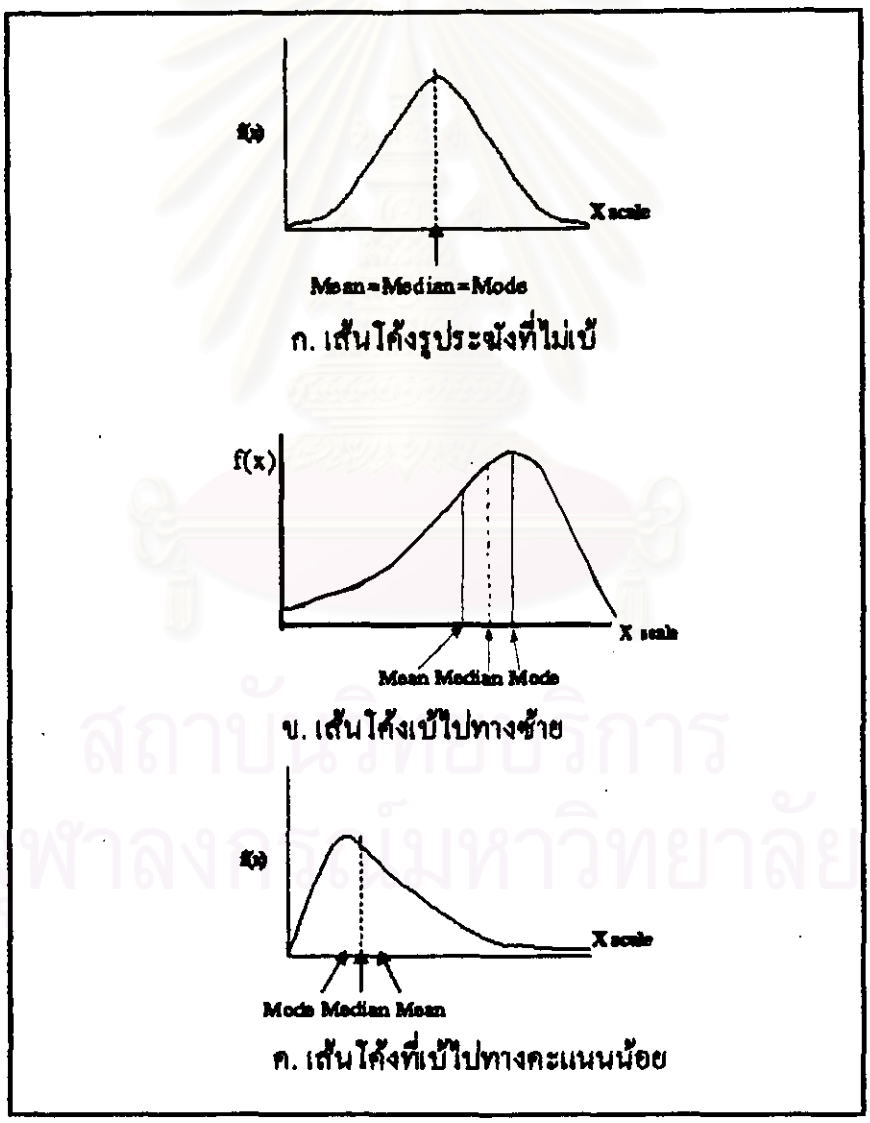
3. การแจกแจงของฟังก์ชันที่ได้จากตัวอย่างสุ่ม ในงานวิจัยนี้จะพิจารณาการแจกแจงของค่าเฉลี่ยตัวอย่าง เมื่อประชากรมีการแจกแจงปกติ และประชากรมีการแจกแจงที่ไม่ใช่การแจกแจงปกติ

4. การทดสอบเทียบความกลมกลืนกัน โดยใช้การทดสอบของโคลโมโกรอฟ-สมินอฟ

2.1 ความเบ้ (Skewness)

ประชากรที่มีการแจกแจงสมมาตรนั้น เส้นโค้งที่ได้จากการแจกแจงจะมีลักษณะเป็นรูประฆังที่สมมาตรกันที่ค่าเฉลี่ย เส้นโค้งทางด้านขวาของค่าเฉลี่ย และทางด้านซ้ายของค่าเฉลี่ย จะมีลักษณะเหมือนกันทุกประการ ค่าเฉลี่ย (Mean) มัธยฐาน (Median) และฐานนิยม (Mode) จะมีค่าเท่ากันหรือทับกันสนิท

แต่ถ้าประชากรมีการแจกแจงไม่สมมาตร มีลักษณะเบ้ไปข้างใดข้างหนึ่ง ค่าเฉลี่ย มัธยฐาน และฐานนิยมจะมีค่าต่างกัน พิจารณาดูรูปต่อไปนี้



รูปที่ 2.1 แสดงเส้นโค้งของการแจกแจงที่ไม่มีความเบ้ เบ้ซ้าย และเบ้ขวา

จะเห็นว่าในรูป ก. ประชากรมีการแจกแจงสมมาตร ซึ่งได้ว่าค่าเฉลี่ย มัธยฐาน และฐานนิยม จะมีค่าเท่ากัน ส่วนในรูป ข. ประชากรมีการแจกแจงเบ้ไปทางซ้าย เพราะพื้นที่ภายใต้เส้นโค้งทางด้านซ้ายของค่าฐานนิยมมีมากกว่าพื้นที่ทางด้านขวาของค่าฐานนิยม และในรูป ค. ประชากรมีการแจกแจงเบ้ไปทางขวา เพราะพื้นที่เส้นโค้งทางด้านขวาของค่าฐานนิยมมากกว่าพื้นที่ทางด้ายซ้ายของค่าฐานนิยม

<u>การวัดความเบ้</u> (Measure of Skewness) จะใช้การวัดความเบ้โดยวิธีโมเมนต์ (Moment)

สูตรสำหรับหาค่าวัดสมมาตร หรือสัมประสิทธ์ความเบ้จากค่าประชากร เป็นดังนี้

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

โดยที่

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

σ คือ ค่าเบี่ยงเบนมาตรฐานของประชากร $\sqrt{V(X)}$

สามารถประมาณค่าวัดสมมาตร หรือสัมประสิทธิ์ความเบ้ จากข้อมูลตัวอย่างได้ค่าสถิติดังนี้

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

โดยที่ $$m_3 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^3}{n}$$

$$m_2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}$$

การวัดความเบ้ด้วยโมเมนต์ศูนย์กลางอันดับที่ 3 จะให้ค่าต่าง ๆ กันดังนี้

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

2.2. <u>ความโด่ง</u> (Kurtosis)

การวัดความโด่ง (Measure of Kurtosis) ก็คือการวัดเส้นโค้งว่าจะมีความโด่งมากน้อยเพียงไร เส้นโค้งที่เราเรียกว่าเส้นโค้งปกติ เส้นโค้งใดที่โด่งผิดจากเส้นโค้งปกติ ก็นับเป็นเส้นโค้งที่ไม่ปกติทั้งสิ้น แม้แต่จะเป็นรูประฆังที่สมมาตรก็ตาม

ความโด่งของการแจกแจงของประชากร มี 3 ลักษณะดังนี้

1. เส้นโค้งที่มีความโด่งเป็นปกติ เรียกว่า เส้นโค้งชนิด Mesokurtic
2. เส้นโค้งที่แบนราบกว่าปกติ เรียกว่า เส้นโค้งชนิด Platykurtic
3. เส้นโค้งที่โด่งกว่าปกติ เรียกว่า เส้นโค้งชนิด Leptokurtic

การวัดความโด่ง หรือการหาสัมประสิทธิ์ความโด่ง (α_4) ซึ่งวัดได้จากค่าโมเมนต์ศูนย์กลางอันดับที่ 4 หารด้วยค่าเบี่ยงเบนมาตรฐานยกกำลังสี่ ดังสูตรต่อไปนี้

$$\alpha_4 = \frac{\mu_4}{\sigma^4} = \frac{E((X-\mu)^4)}{(V(X))^2}$$

โดยที่

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

σ คือ ค่าเบี่ยงเบนมาตรฐานของประชากร $\sqrt{V(X)}$

สามารถประมาณค่าวัดความโด่ง หรือสัมประสิทธิ์ความโด่งจากข้อมูลตัวอย่างได้ค่าสถิติดังนี้

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

โดยที่ $$m_4 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^4}{n}$$

$$m_2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}$$

การวัดความโด่งโดยใช้โมเมนต์ศูนย์กลางอันดับที่ 4 จะให้ค่าต่าง ๆ กันดังนี้

1. ถ้า α_4 เท่ากับ 3 แสดงว่าเส้นโค้งมีความโด่งเป็นปกติ (Mesokurtic)
2. ถ้า α_4 น้อยกว่า 3 แสดงว่าเส้นโค้งแบนราบกว่าปกติ (Platykurtic)
3. ถ้า α_4 มากกว่า 3 แสดงว่าเส้นโค้งที่โด่งกว่าปกติ (Leptokurtic)

ในงานวิจัยครั้งนี้ การพิจารณาเลือกค่าขนาดตัวอย่าง n สำหรับประมาณการแจกแจงของตัวสถิติทดสอบที เมื่อประชากรมีการแจกแจงแบบต่างๆ ซึ่งต้องพิจารณาก่อนว่าประชากรนั้นมีความเบ้ ความโด่งเป็นเท่าไร โดยวัดจากค่าสัมประสิทธิ์ความเบ้และความโด่งจากค่าประชากร แต่ถ้ากรณีไม่ทราบค่าวัดของประชากร สามารถประมาณค่าวัดเหล่านี้ได้จากค่าตัวอย่าง

2.3 การแจกแจงที่เกี่ยวข้อง

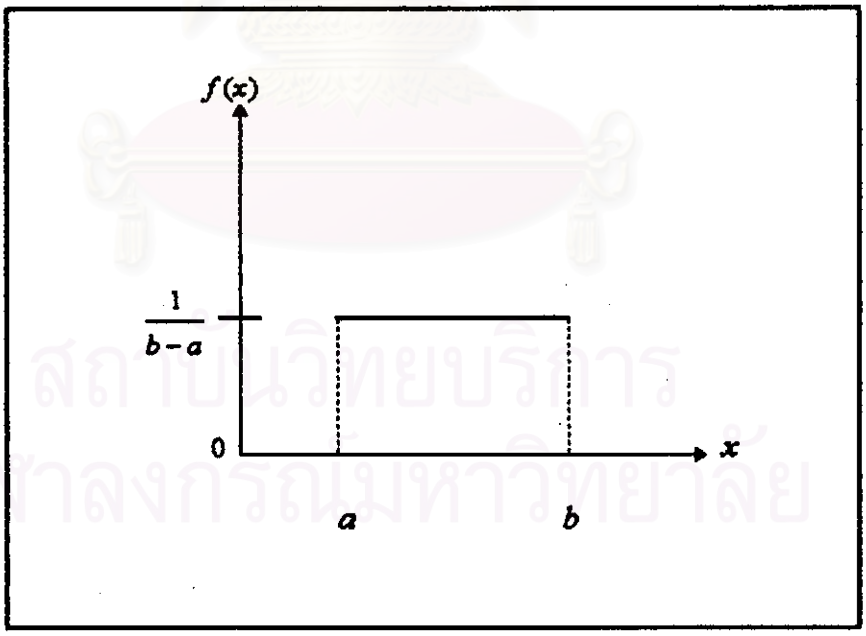
2.3.1 การแจกแจงเอกรูป (Uniform Distribution)

ให้ X เป็นตัวแปรสุ่มแบบต่อเนื่อง ซึ่งมีค่าได้ทุกค่าจริงในช่วงใดช่วงหนึ่ง ด้วยความน่าจะเป็นเท่า ๆ กัน จะเรียกว่า ตัวแปรสุ่ม X มีการแจกแจงเอกรูป และมีฟังก์ชันความหนาแน่นคือ

$$f(x;a,b)=\frac{1}{(b-a)} \qquad ; \; a \le x \le b$$

โดยที่ a และ b เป็นค่าคงที่ และ $a < b$

จะพบว่าค่าของ X เป็นไปได้ทุกค่าในช่วง (a,b) ดังแสดงในรูปที่ 2.2



รูปที่ 2.2 แสดงฟังก์ชันความหนาแน่นของการแจกแจงเอกรูป

ลักษณะทั่วไปของตัวแปรสุ่ม X ที่มีการแจกแจงเอกรูป

1. ค่าเฉลี่ยของตัวแปรสุ่ม X

$$E(X) = \mu = \frac{a+b}{2}$$

2. ค่าความแปรปรวนของตัวแปรสุ่ม X

$$V(X) = \sigma^2 = \frac{(b-a)^2}{12}$$

3. สัมประสิทธิ์ความเบ้ (Coefficient of skewness) เท่ากับ 0
4. สัมประสิทธิ์ความโด่ง (Coefficient of kurtosis) เท่ากับ 1.8

2.3.2 <u>การแจกแจงโลจิสติก</u> (Logistic Distribution)

ให้ X เป็นตัวแปรสุ่มแบบต่อเนื่อง X มีการแจกแจงความน่าจะเป็นโลจิสติกด้วยพารามิเตอร์ a และ b ถ้า X มีฟังก์ชันความหนาแน่นอยู่ในรูป

$$f(x;a,b) = \frac{\exp[(x-a)/b]}{b\{1+\exp[(x-a)/b]\}^2} \quad ; -\infty < x < \infty \text{ , } b > 0$$

$$; \ -\infty < a < \infty$$

โดยที่ a เป็นพารามิเตอร์ตำแหน่ง (location parameter)

b เป็นพารามิเตอร์มาตรส่วน (scale parameter)

ลักษณะทั่วไปของตัวแปรสุ่ม X ที่มีการแจกแจงโลจิสติค

1. ค่าเฉลี่ยของตัวแปรสุ่ม X

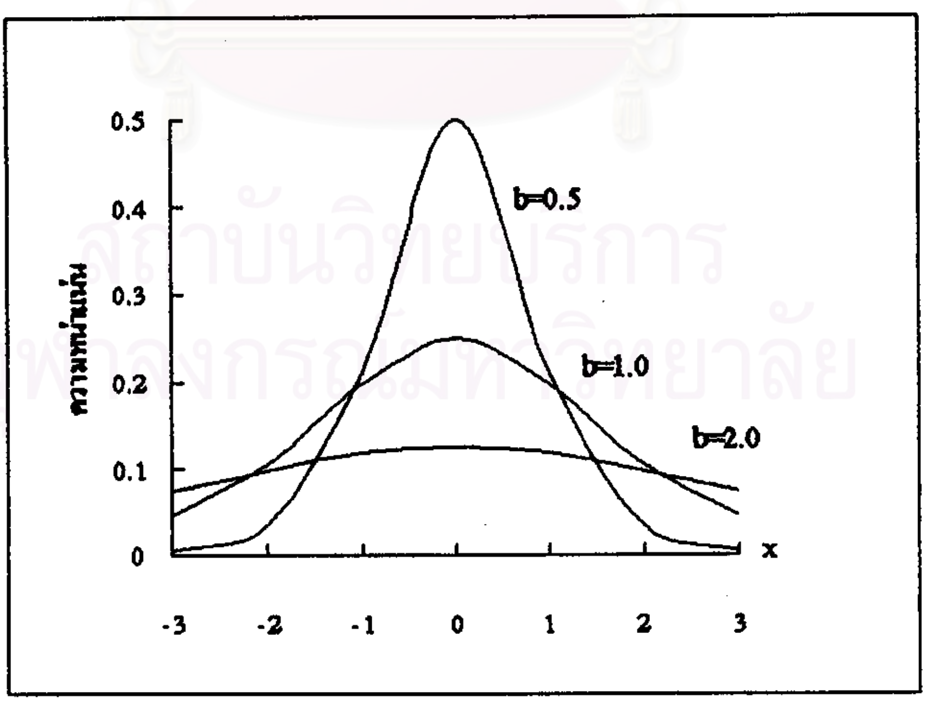
$$E(X) = \mu = a$$

2. ค่าความแปรปรวนของตัวแปรสุ่ม X

$$V(X) = \sigma^2 = \pi^2 b^2/3$$

3. สัมประสิทธิ์ความเบ้ (Coefficient of skewness) เท่ากับ 0
4. สัมประสิทธิ์ความโด่ง (Coefficient of kurtosis) เท่ากับ 4.2

ฟังก์ชันความหนาแน่นของการแจกแจงโลจิสติค มีลักษณะสมมาตรคล้ายการแจกแจงปกติ แต่ลักษณะที่ส่วนปลายโค้งของการแจกแจงหนา ซึ่งกราฟของฟังก์ชันความหนาแน่นของการแจกแจงจะขึ้นอยู่กับพารามิเตอร์ b คือถ้า b มีค่ามากปลายโค้งก็จะหนากว่า กราฟแสดงฟังก์ชันความหนาแน่นของการแจกแจงโลจิสติคแสดงดังรูปที่ 2.3



รูปที่ 2.3 แสดงฟังก์ชันความหนาแน่นของการแจกแจงโลจิสติค

2.3.3 การแจกแจงที (t Distribution)

ถ้าให้ Z เป็นตัวแปรสุ่มแบบปกติมาตรฐาน และ Y เป็นตัวแปรสุ่มแบบไคกำลังสอง ด้วยระดับขั้นความเสรี ν และถ้า Z และ Y เป็นตัวแปรสุ่มอิสระกันแล้ว ตัวแปรสุ่ม $T=\dfrac{Z}{\sqrt{Y/\nu}}$ จะมีการแจกแจงที ด้วยระดับขั้นความเสรีเท่ากับ ν ซึ่งมีฟังก์ชันความหนาแน่นอยู่ในรูป

$$f(t,\nu)=\frac{\Gamma[(\nu+1)/2]}{(\pi\nu)^{1/2}\Gamma(\nu/2)[1+(t^2/\nu)]^{(\nu+1)/2}} \qquad , -\infty<t<\infty$$

โดยที่ ν เป็นพารามิเตอร์สัณฐาน (shape parameter)

ลักษณะทั่วไปของตัวแปรสุ่ม T ที่มีการแจกแจงที

1. ค่าเฉลี่ยของตัวแปรสุ่ม T

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

2. ค่าความแปรปรวนของตัวแปรสุ่ม T

$$V(T)=\sigma^2=\frac{\nu}{(\nu-2)} \qquad ; \ \nu>2$$

3. สัมประสิทธิ์ความเบ้ (Coefficient of skewness) เท่ากับ 0
4. สัมประสิทธิ์ความโด่ง (Coefficient of kurtosis)

$$\alpha_4=\frac{3(\nu-2)}{(\nu-4)} \qquad ; \ \nu>4$$

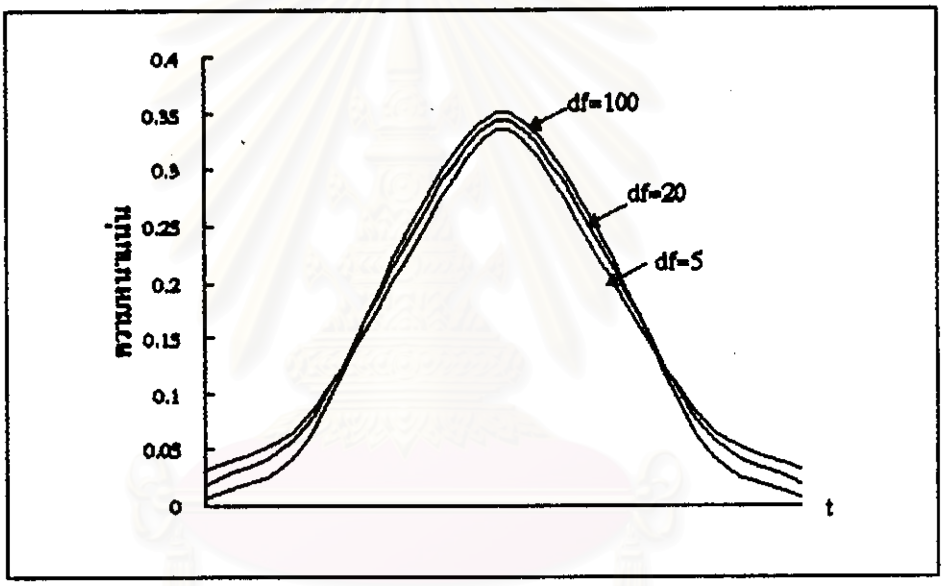
ลักษณะของเส้นโค้งที

1. ตัวแปร t มีค่าระหว่าง $-\infty$ และ ∞
2. มีค่าเฉลี่ยเท่ากับศูนย์สำหรับ $\nu>1$

3. มีความสมมาตรรอบค่าเฉลี่ย ถ้าระดับขั้นความเสรีน้อยจะทำให้เส้นโค้งที่แบนราบกว่ากรณีที่ระดับขั้นความเสรีมาก ซึ่งทำให้เส้นโค้งที่แบนราบกว่าเส้นโค้งปกติมาตรฐาน

4. ค่าความแปรปรวนมากกว่า 1

เส้นโค้งทีจะคล้ายกับเส้นโค้งปกติมาตรฐาน คือ เป็นระฆังคว่ำ แต่เส้นโค้งที จะลาดกว่าเส้นโค้งปกติมาตรฐาน การที่เส้นโค้งทีจะโด่งหรือไม่ขึ้นกับค่าระดับขั้นความเสรี แสดงดังรูปที่ 2.4



รูปที่ 2.4 แสดงฟังก์ชันความหนาแน่นของการแจกแจงที

2.3.4 การแจกแจงไคกำลังสอง (Chi-square Distribution)

ถ้า X เป็นตัวแปรสุ่มที่มีการแจกแจงปกติที่มีค่าเฉลี่ย μ และค่าความแปรปรวน σ^2 จะได้ $Z = \dfrac{X-\mu}{\sigma}$ มีการแจกแจงปกติมาตรฐาน และ $Z^2 = \dfrac{(X-\mu)^2}{\sigma^2}$ มีการแจกแจงไคกำลังสอง ด้วยระดับขั้นความเสรีเท่ากับ 1

ถ้า $X_1, X_2, \ldots, X_n$ เป็นตัวอย่างสุ่มขนาด n จากการแจกแจงปกติที่มีค่าเฉลี่ย μ และความแปรปรวน σ^2 จะได้ว่า $\dfrac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma^2}$ จะมีการแจกแจงไคกำลังสองด้วยระดับขั้นความเสรีเท่ากับ n หรือเขียนได้ว่า

$$\chi_n^2 = \sum_{i=1}^{n} Z^2 = \frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma^2}$$

ดังนั้นถ้า X เป็นตัวแปรสุ่มแบบไคกำลังสองด้วยระดับขั้นความเสรี ν ฟังก์ชันความหนาแน่นของ X คือ

$$f(x;\nu) = \frac{x^{(\nu-2)/2}\exp(-x/2)}{2^{\nu/2}\Gamma(\nu/2)} \quad ; \; 0 \le x < \infty$$

ลักษณะทั่วไปของตัวแปรสุ่ม X ที่มีการแจกแจงไคกำลังสอง

1. ค่าเฉลี่ยของตัวแปรสุ่ม X

$$E(X) = \mu = \nu$$

2. ค่าความแปรปรวนของตัวแปรสุ่ม X

$$V(X) = \sigma^2 = 2\nu$$

3. สัมประสิทธิ์ความเบ้ (Coefficient of skewness) เป็นดังนี้

$$\alpha_3 = 2^{3/2}\,\nu^{-1/2}$$

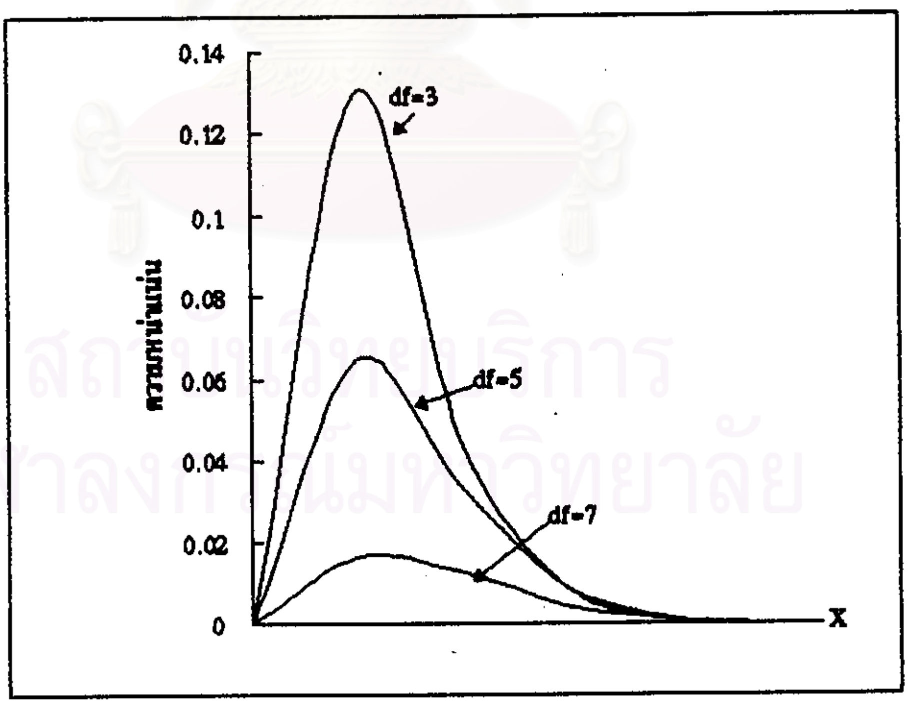
4. สัมประสิทธิ์ความโด่ง (Coefficient of kurtosis) เป็นดังนี้

$$\alpha_4 = 3 + 12/\nu$$

ลักษณะของเส้นโค้ง χ_n^2 คือ

1. ค่าตัวแปร χ_n^2 มีค่าตั้งแต่ 0 ถึง ∞
2. เส้นโค้งมีลักษณะเบ้ขวา ซึ่งการเบ้จะแตกต่างกันโดยขึ้นอยู่กับระดับขั้นความเสรี (ν)
3. ถ้าระดับขั้นความเสรีมีค่ามาก เส้นโค้งไคกำลังสองจะคล้ายเส้นโค้งปกติ

ฟังก์ชันความหนาแน่นของการแจกแจงไคกำลังสอง แสดงดังรูปที่ 2.5



รูปที่ 2.5 แสดงฟังก์ชันความหนาแน่นของการแจกแจงไคกำลังสอง

2.3.5 การแจกแจงลอกนอร์มัล (Lognomal Distribution)

ตัวแปรสุ่ม X มีการแจกแจงความน่าจะเป็นแบบลอกนอร์มัลด้วยพารามิเตอร์ μ และ σ^2 ถ้า X มีฟังก์ชันความหนาแน่น

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

โดยที่ $\exp(\sigma^2)$ เป็นพารามิเตอร์มาตรส่วน (scale parameter)

μ เป็นพารามิเตอร์สัณฐาน (shape parameter)

ลักษณะทั่วไปของตัวแปรสุ่ม X ที่มีการแจกแจงลอกนอร์มัล

1. ค่าเฉลี่ยของตัวแปรสุ่ม X

$$E(X)=\mu=\exp(\mu+\sigma^2/2)$$

2. ค่าความแปรปรวนของตัวแปรสุ่ม X

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

โดยที่ $m=\exp(\mu)$

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

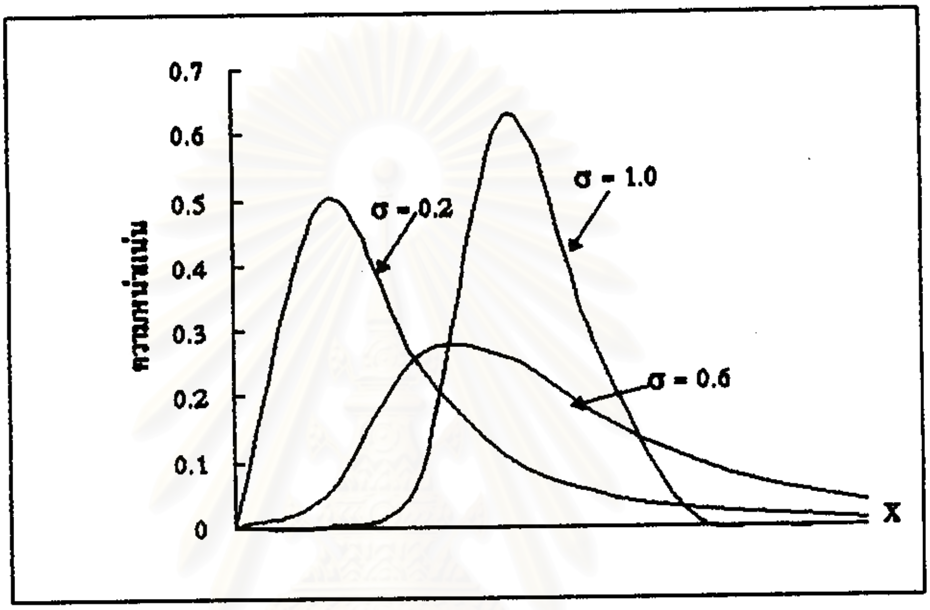
3. สัมประสิทธิ์ความเบ้ (Coefficient of skewness) เป็นดังนี้

$$\alpha_3=(\omega+2)(\omega-1)^{1/2}$$

4. สัมประสิทธิ์ความโด่ง (Coefficient of kurtosis) เป็นดังนี้

$$\alpha_4=\omega^4+2\omega^3+3\omega^2-3$$

กราฟแสดงฟังก์ชันความหนาแน่นของการแจกแจงลอกนอร์มัล มีลักษณะเบ้ขวา และถ้าค่าสัมประสิทธิ์ความแปรผันต่ำ ๆ หรือค่าความแปรปรวนต่ำๆ จะทำให้กราฟฟังก์ชันความหนาแน่นเข้าใกล้การแจกแจงปกติมากขึ้น ดังแสดงในรูปที่ 2.6



รูปที่ 2.6 แสดงฟังก์ชันความหนาแน่นของการแจกแจงลอกนอร์มัล

2.3.6 การแจกแจงแลมดาของตูกีร์ (Tukey's Lamda Distribution)

Ramberg และ Schmeiser ได้เสนอวิธีการสร้างตัวแปรสุ่มที่ขึ้นอยู่กับความเบ้ (skewness : α_3) และความโด่ง (kurtosis : α_4) โดยตัวแปรสุ่มนี้มีการแจกแจงที่เรียกว่า "การแจกแจงแลมดาของตูกีร์" โดยที่ตัวแปรสุ่มนั้นจะถูกกำหนดจากค่าพารามิเตอร์ 4 ค่า ซึ่งสัมพันธ์กับค่าความเบ้และความโด่ง ดังนี้

$$X = R(p) = \lambda_1 + \left[p^{\lambda_3} - (1-p)^{\lambda_4}\right]/\lambda_2 \qquad ; 0 \le p \le 1 \qquad (2.1)$$

โดยที่ p เป็นเลขสุ่มที่มีค่าระหว่าง 0 และ 1

λ_1 เป็นพารามิเตอร์กำหนดตำแหน่ง (location parameter)

λ_2 เป็นพารามิเตอร์มาตราส่วน (scale parameter)

λ_3, λ_4 เป็นพารามิเตอร์สัณฐาน (shape parameter) ซึ่งขึ้นกับค่าความเบ้และความโด่งที่กำหนด ถ้าการแจกแจงเป็นแบบสมมาตร จะได้ว่า $\lambda_3 = \lambda_4$

ดังนั้นฟังก์ชันความหนาแน่นของตัวแปรสุ่ม X ที่ได้จากสมการ (2.1) คือ

$$\begin{aligned} f(x) &= f[R(p)] \\ &= 1/R'(p) \qquad ; \quad 0 \le p \le 1 \qquad (2.2) \\ &= \lambda_2[\lambda_3 p^{\lambda_3 - 1} + \lambda_4 (1-p)^{\lambda_4 - 1}]^{-1} \end{aligned}$$

โดยที่

$$R'(p) = dR(p)/dp$$

Ramberg และ Schmeiser ได้แสดงค่าโมเมนต์ที่ k เมื่อ $\lambda_1 = 0$ ได้ดังสมการ

$$E(X^k) = \lambda_2^{-k} \sum_{i=0}^{k} \binom{k}{i} (-1)^i \beta(\lambda_3(k-i)+1, \lambda_4 i+1) \qquad (2.3)$$

โดยที่ β คือค่าเบต้าฟังก์ชัน (beta function)

จากสมการที่ (2.3) โมเมนต์ที่ k จะหาค่าไม่ได้ เมื่อค่าเบต้าฟังก์ชันมีค่าเป็นลบ ดังนั้นโมเมนต์ที่ k จะหาค่าได้ก็ต่อเมื่อ $-1/k < \min(\lambda_3, \lambda_4)$

ดังนั้นจากสมการที่ (2.3) สามารถหาค่าเฉลี่ย, ความแปรปรวน,โมเมนต์ที่ 3 รอบค่าเฉลี่ย ($\mu_3 = E(X-\mu)^3$) และโมเมนต์ที่ 4 รอบค่าเฉลี่ย ($\mu_4 = E(X-\mu)^4$) จากการแจกแจงได้ดังนี้

$$\begin{aligned} \mu &= \lambda_1 + A/\lambda_2 \\ \sigma^2 &= (B - A^2)/\lambda_2^2 \\ \mu_3 &= (C - 3AB + 2A^3)/\lambda_2^3 \\ \mu_4 &= (D - 4AC + 6A^2B - 3A^4)/\lambda_2^4 \end{aligned}$$

โดยที่

$$A = 1/(1+\lambda_3) - 1/(1+\lambda_4)$$

$$B = 1/(1+2\lambda_3) + 1/(1+2\lambda_4) - 2\beta(1+\lambda_3, 1+\lambda_4)$$

$$C = 1/(1+3\lambda_3) - 3\beta(1+2\lambda_3, 1+\lambda_4) + 3\beta(1+\lambda_3, 1+2\lambda_4) - 1/(1+3\lambda_4)$$

$$D = 1/(1+4\lambda_3) - 4\beta(1+3\lambda_3, 1+\lambda_4) + 6\beta(1+2\lambda_3, 1+2\lambda_4) - 4\beta(1+\lambda_3, 1+3\lambda_4) + 1/(1+4\lambda_4)$$

ดังนั้นค่าความเบ้และความโด่ง เป็นไปตามสมการดังนี้

$$\alpha_3 = \mu_3/\sigma^3 \tag{2.4}$$

และ

$$\alpha_4 = \mu_4/\sigma^4 \tag{2.5}$$

เราสามารถหาค่า $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ เมื่อกำหนดค่าความเบ้และความโด่งต่าง ๆ ได้จากตาราง Ramberg แสดงในภาคผนวก โดยที่ค่า λ_1, λ_2 เป็นค่าที่ค่าเฉลี่ยเป็นศูนย์และความแปรปรวนเท่ากับ 1 แต่ถ้าค่าเฉลี่ยเป็น μ และความแปรปรวนเป็น σ^2 จะต้องแปลงค่า λ_1, λ_2 จากตารางดังนี้

$$\lambda_1(\mu, \sigma^2) = \lambda_1(0,1)\sigma + \mu \tag{2.6}$$

$$\lambda_2(\mu, \sigma^2) = \lambda_2(0,1)/\sigma \tag{2.7}$$

ค่าประมาณความเบ้ $\hat{\alpha}_3$, ความโด่ง $\hat{\alpha}_4$ จากข้อมูลตัวอย่างได้ดังนี้

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

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

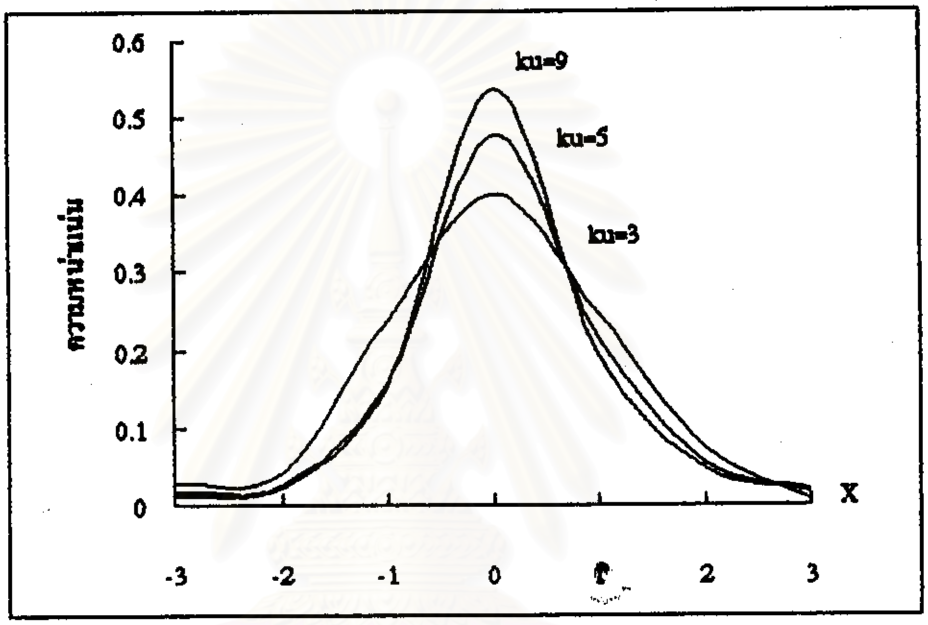
โดยที่

$$m_2 = \sum_{i=1}^{n}(x_i - \bar{x})^2/n$$

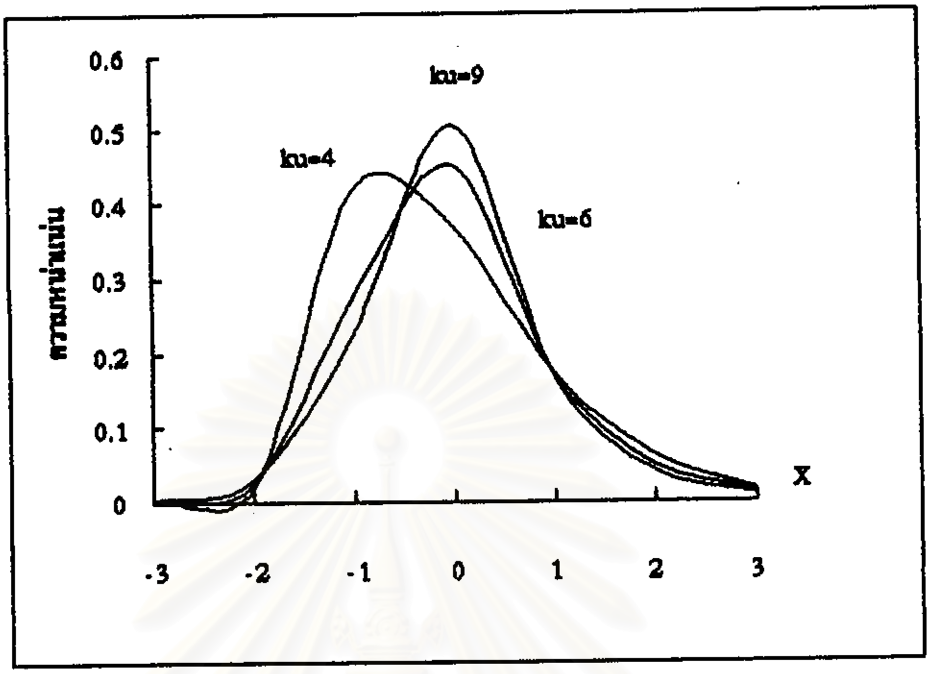
$$m_3 = \sum_{i=1}^{n}(x_i - \bar{x})^3/n$$

และ $$m_4 = \sum_{i=1}^{n} (x_i - \bar{x})^4 / n$$

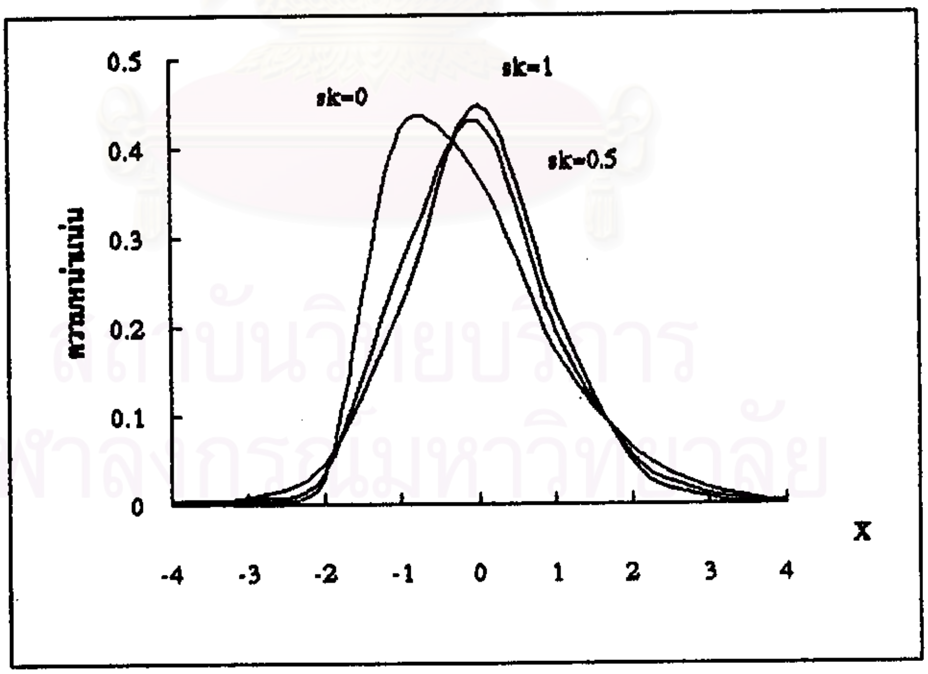
กราฟแสดงฟังก์ชันความหนาแน่นของการแจกแจงแลมดาของตูกีร์ ที่ค่าความเบ้ และความโด่งต่าง ๆ แสดงดังรูปที่ 2.7-2.9



รูปที่ 2.7 แสดงฟังก์ชันความหนาแน่นของการแจกแจงแลมดาของตูกีร์ ที่ความเบ้เท่ากับ 0 , ความโด่งเท่ากับ 3,5,9



<u>รูปที่ 2.8</u> แสดงฟังก์ชันความหนาแน่นของการแจกแจงแลมดาของตูกีร์ ที่ความเบ้เท่ากับ 1 , ความโด่งเท่ากับ 4,6,9



<u>รูปที่ 2.9</u> แสดงฟังก์ชันความหนาแน่นของการแจกแจงแลมดาของตูกีร์ ที่ความเบ้เท่ากับ 0,0.5,1 , ความโด่งเท่ากับ 4

2.4 การแจกแจงของฟังก์ชันที่ได้จากตัวอย่างสุ่ม

2.4.1 ค่าคาดหวังและค่าความคลาดเคลื่อนมาตรฐานของฟังก์ชันที่ได้จากตัวอย่างสุ่ม

เนื่องจากฟังก์ชันที่ได้จากตัวอย่างสุ่มเป็นตัวแปรสุ่มที่มีการแจกแจงความน่าจะเป็น จึงสามารถที่จะหาค่าคาดหวัง และค่าความคลาดเคลื่อนมาตรฐาน (Standard Error) ของฟังก์ชันหรือตัวแปรสุ่มนี้ได้ สมมติว่า จากตัวอย่างสุ่ม $(Y_1,Y_2,\ldots,Y_n)$ สร้างฟังก์ชัน $W=\bar{Y}$ จะมีการแจกแจงความน่าจะเป็น โดยค่าคาดหวังของ W หรือ $E(W)$ คือค่าเฉลี่ยของการแจกแจงของ W และค่าความคลาดเคลื่อนมาตรฐานของ W หรือ $S.E(W)$ คือรากที่สองของค่าความแปรปรวนของ W

ยกตัวอย่างเช่น กรณีที่ฟังก์ชันที่ได้จากตัวอย่างสุ่มเป็นค่าเฉลี่ยตัวอย่าง ให้ $\bar{Y}$ เป็นค่าเฉลี่ยตัวอย่างของตัวอย่างสุ่ม $(Y_1,Y_2,\ldots,Y_n)$ ซึ่งเลือกมาจากประชากรที่มีค่าเฉลี่ย μ และความแปรปรวน σ^2 จะได้ว่าค่าเฉลี่ยตัวอย่าง $\bar{Y}$ จะมีค่าคาดหวังเช่นเดียวกับค่าคาดหวังของประชากร Y และจะมีค่าความแปรปรวนเท่ากับค่าความแปรปรวนของ Y หารด้วยขนาดตัวอย่าง กล่าวคือ

$$E(\bar{Y}) = E(Y) = \mu$$

$$V(\bar{Y}) = \frac{V(Y)}{n} = \frac{\sigma^2}{n}$$

และ $$S.E.(\bar{Y}) = \frac{\sigma}{\sqrt{n}}$$

2.4.2 การแจกแจงของค่าเฉลี่ยตัวอย่าง

สมมติว่าตัวอย่างสุ่ม $(Y_1,Y_2,\ldots,Y_n)$ ถูกเลือกจากประชากรที่มีการแจกแจงแบบปกติ ที่มีค่าเฉลี่ย μ และความแปรปรวน σ^2 กล่าวคือ $Y_1,Y_2,\ldots,Y_n$ ต่างเป็นตัวแปรสุ่มอิสระที่มีการแจกแจงแบบ $N(\mu,\sigma^2)$ เดียวกัน ฟังก์ชันของตัวอย่างสุ่มที่ต้องการจะหาการแจกแจงในที่นี้ คือ ค่าเฉลี่ยตัวอย่าง $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$ การหาการแจกแจงของค่าเฉลี่ยตัวอย่าง ก็คือ การหาว่าค่าเฉลี่ยตัวอย่างมีการแจกแจงรูปแบบใด และค่าพารามิเตอร์ ซึ่งในที่นี้คือ μ และ

σ^2 เท่ากับเท่าไร สำหรับรูปแบบการแจกแจงของ $\bar{Y}$ อาจแสดงได้โดยอาศัยวิธีการ เช่น คุณสมบัติของฟังก์ชันที่ให้โมเมนต์ (moment generation funotion) ว่าการแจกแจงของ $\bar{Y}$ ยังคงเป็นแบบปกติเช่นเดียวกับ Y_i แต่มีพารามิเตอร์บางค่าที่เปลี่ยนไปทั้งนี้จะเห็นได้ว่า

$$E(\bar{Y}) = E\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right)$$

$$= \frac{1}{n}\sum^{n} E(Y_i) = \frac{1}{n}\sum^{n} \mu$$

$$= \frac{1}{n}(n\mu) = \mu$$

และ $$V(\bar{Y}) = V\left(\frac{1}{n}\sum^{n} Y_i\right) = \frac{1}{n^2}\sum^{n} Y_i$$

$$= \frac{1}{n^2}\sum^{n} \sigma^2$$

$$= \frac{\sigma^2}{n}$$

2.4.3 <u>การแจกแจงของค่าเฉลี่ยตัวอย่าง เมื่อไม่ทราบค่าเบี่ยงเบนมาตรฐานของประชากร และประชากรมีการแจกแจงปกติ</u>

จากหัวข้อ 2.4.2 ทราบว่าถ้าเลือกตัวอย่างจากประชากรที่มีค่าเฉลี่ย μ และความแปรปรวน σ^2 ค่าเฉลี่ยตัวอย่าง $\bar{Y}$ จะมีการแจกแจงเป็นปกติ โดยมีค่าเฉลี่ย μ และความแปรปรวน $\frac{\sigma^2}{n}$ ทำให้ $\frac{(\bar{Y}-\mu)}{\sigma/n}$ มีการแจกแจงแบบ $N(0,1)$ อย่างไรก็ตาม ในทางปฏิบัติ บ่อยครั้งเราจะไม่ทราบค่าเบี่ยงเบนมาตรฐานของประชากร ในกรณีเช่นนี้ มักจะประมาณ σ ด้วยค่าเบี่ยงเบนมาตรฐานของตัวอย่าง S ตัวสถิติที่น่าสนใจจึงเป็น $\frac{(\bar{Y}-\mu)}{S/\sqrt{n}}$ ซึ่งไม่ได้มีการแจกแจงปกติอีกต่อไป แต่เป็นตัวสถิติที่เป็นฟังก์ชันของตัวแปรสุ่ม 2 ตัว คือ $\bar{Y}$ และ S

ถ้า Z มีการแจกแจงแบบ $N(0,1)$ และ U มีการแจกแจงไคกำลังสองที่ระดับขั้นความเสรี v โดยที่ Z และ U เป็นอิสระต่อกันแล้ว $Z/\sqrt{U/v}$ จะมีการแจกแจงที ด้วยระดับขั้นความเสรี v สำหรับตัวสถิติที่ต้องการหาการแจกแจงในที่นี้ จะเห็นว่าถ้าให้

$$Z = \frac{\bar{Y} - \mu}{\sigma/\sqrt{n}}$$

Z มีการแจกแจงแบบ $N(0,1)$

และให้ $$U = \frac{(n-1)S^2}{\sigma^2}$$

โดยที่

$$S^2 = \frac{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}{n-1}$$

U จะมีการแจกแจงไคกำลังสองด้วยระดับขั้นความเสรีเท่ากับ $n-1$

ดังนั้น

$$\frac{\bar{Y} - \mu}{S/\sqrt{n}} = \frac{(\bar{Y} - \mu)/(\sigma/\sqrt{n})}{(S/\sqrt{n})/(\sigma/\sqrt{n})}$$

$$= \frac{Z}{S/\sigma}$$

$$= \frac{Z}{\sqrt{((n-1)S^2)/(\sigma^2(n-1))}}$$

$$= \frac{Z}{\sqrt{U/(n-1)}}$$

โดยที่ Z และ U เป็นอิสระต่อกัน

ดังนั้นตัวสถิติทดสอบ

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

จึงมีการแจกแจงที ด้วยระดับขั้นความเสรีเท่ากับ $n-1$

ตัวสถิติทดสอบทีดังกล่าวมีความสำคัญมากในการประมาณค่าแบบช่วง และการทดสอบสมมติฐานเกี่ยวกับค่าเฉลี่ยประชากร กรณีไม่ทราบค่าเบี่ยงเบนมาตรฐานของประชากร

2.4.4 <u>การแจกแจงของค่าเฉลี่ยตัวอย่าง เมื่อประชากรมีการแจกแจงแบบอื่นๆ ที่ไม่ใช่การแจกแจงปกติ</u>

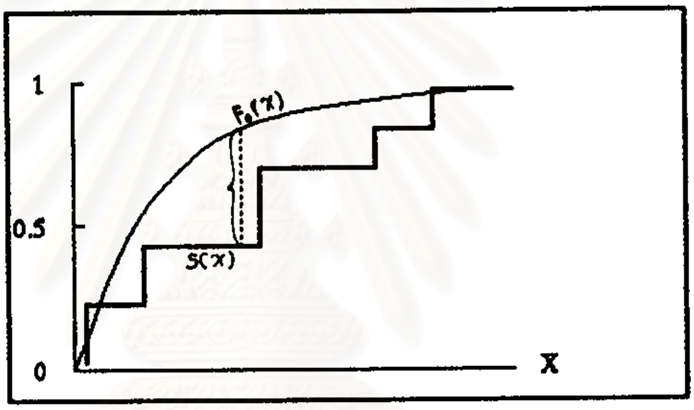
ในบางครั้งประชากรที่เราสนใจอาจมีการแจกแจงแบบอื่น ๆ ที่ไม่ใช่การแจกแจงปกติ โดยปกติเราจะอาศัยทฤษฎีบทลิมิตเข้าสู่ส่วนกลาง (Central Limit Theorm) เพื่อประมาณการแจกแจงของค่าเฉลี่ยตัวอย่างดังนี้

ถ้าสุ่มตัวอย่างขนาด n $(Y_1,Y_2,\ldots,Y_n)$ จากประชากรใด ๆ ที่มีค่าเฉลี่ย μ และความแปรปรวน σ^2 แล้ว ถ้าตัวอย่างมีขนาดใหญ่ ค่าเฉลี่ยตัวอย่าง $\bar{Y}$ จะมีการแจกแจงเข้าสู่การแจกแจงปกติที่มีค่าเฉลี่ย μ และความแปรปรวน σ^2/n หรือ $\dfrac{\bar{Y}-\mu}{\sigma/\sqrt{n}}$ จะมีการแจกแจงโดยประมาณปกติมาตรฐาน แต่ถ้าไม่ทราบค่าเบี่ยงเบนมาตรฐานของประชากร(σ) จะประมาณค่าส่วนเบี่ยงเบนมาตรฐานของประชากรด้วยค่าเบี่ยงเบนมาตรฐานของตัวอย่าง (S) ดังนั้นการแจกแจงของ $T=\dfrac{(\bar{Y}-\mu)}{S/\sqrt{n}}$ ซึ่งเรียกว่า "ตัวสถิติทดสอบที" จะมีการแจกแจงที ถ้าขนาดตัวอย่างมีขนาดใหญ่พอ

อย่างไรก็ตาม ขนาดตัวอย่างจะขึ้นอยู่กับลักษณะการแจกแจงของประชากร ถ้าการแจกแจงของประชากรใกล้เคียงการแจกแจงปกติ ขนาดตัวอย่างที่ใช้ก็ไม่จำเป็นต้องใหญ่มาก แต่ถ้าการแจกแจงของประชากรแตกต่างจากการแจกแจงปกติมาก เช่น มีความเบ้มาก ขนาดตัวอย่างก็ควรสูงขึ้น

2.5. <u>การทดสอบเทียบความกลมกลืนกัน</u> (Goodness of fit test)

การทดสอบเทียบความกลมกลืนกัน ในงานวิจัยนี้จะใช้การทดสอบของโคลโมโกรอฟ-สมินอฟ (Kolmogorov-smirnov test) ซึ่งสถิติทดสอบโคโมโกรอฟ-สมินอฟ เป็นทดสอบที่คิดขึ้นโดย A.N. Kolmogorov (1993) และ N.V. Smirnov (1939) ดังนั้นจึงเรียกการทดสอบนี้ว่า การทดสอบโคลโมโกรอฟ-สมินอฟ หลักการของสถิติทดสอบนี้คือ การวัดระยะที่ห่างมากที่สุดระหว่างกราฟของ $S(x)$ และ $F_0(x)$ โดยที่ $S(x)$ แทนฟังก์ชันการแจกแจงความถี่สัมพัทธ์สะสมของตัวอย่าง และ $F_0(x)$ แทนฟังก์ชันการแจกแจงความถี่สะสมภายใต้ H_0 (Hypothesis Distribution Function) ดังแสดงในรูปที่ 2.10



<u>รูปที่ 2.10</u> กราฟเปรียบเทียบฟังก์ชันการแจกแจงความถี่สัมพัทธ์สะสมของตัวอย่างและฟังก์ชันการแจกแจงความถี่สะสมภายใต้ H_0

คำนวณค่าสถิติทดสอบ

$$D = \max(D^+, D^-)$$

โดยที่
$$D^+ = \max[i/m, F(x_i)]$$
$$D^- = \max[F(x_i), (i-1)/m]$$

และ $F(x_i)$ คือ ความถี่สะสมสัมพัทธ์ที่ได้จากทฤษฎี หรือ ฟังก์ชันการแจกแจงสะสม , $i = 1,2,...,m$ (m คือ จำนวนข้อมูลที่นำมาทดสอบ) ในงานวิจัยนี้ $F(x_i)$ ก็คือ ฟังก์ชันการแจกแจงสะสมของการแจกแจงที

i/m ,$(i-1)/m$ คือ $S(x)$ ความถี่สัมพัทธ์สะสมที่ได้จากตัวอย่าง หรือ ความถี่สะสมที่สังเกตได้ในรูปของสัดส่วน (Empirical Distribution Function)

นำค่า D ที่คำนวณได้ไปเปรียบเทียบกับค่าวิกฤตในตารางโคลโมโกรอฟ-สมินอฟ ถ้า D ที่คำนวณได้มากกว่า $D^*_{\alpha,\beta,m}$ จะสรุปได้ว่า การแจกแจงค่าที่ได้จากการทดลองจะไม่เป็นไปตามสมมติฐานที่ตั้งไว้