

บทที่ 2

ทฤษฎีและสถิติที่เกี่ยวข้อง

งานวิจัยนี้มีจุดมุ่งหมายเพื่อศึกษาอำนาจการทดสอบของสถิติทดสอบค่าเฉลี่ยของประชากรกลุ่มเดียว เมื่อประชากรมีการแจกแจงแบบเบ้ขวา โดยใช้การแจกแจงแลมดาของคูร์ ซึ่งเป็นการแจกแจงที่กำหนดโดยค่าความเบ้และความโค้งของข้อมูล สำหรับบทนี้จะกล่าวรายละเอียดตามหัวข้อต่อไปนี้

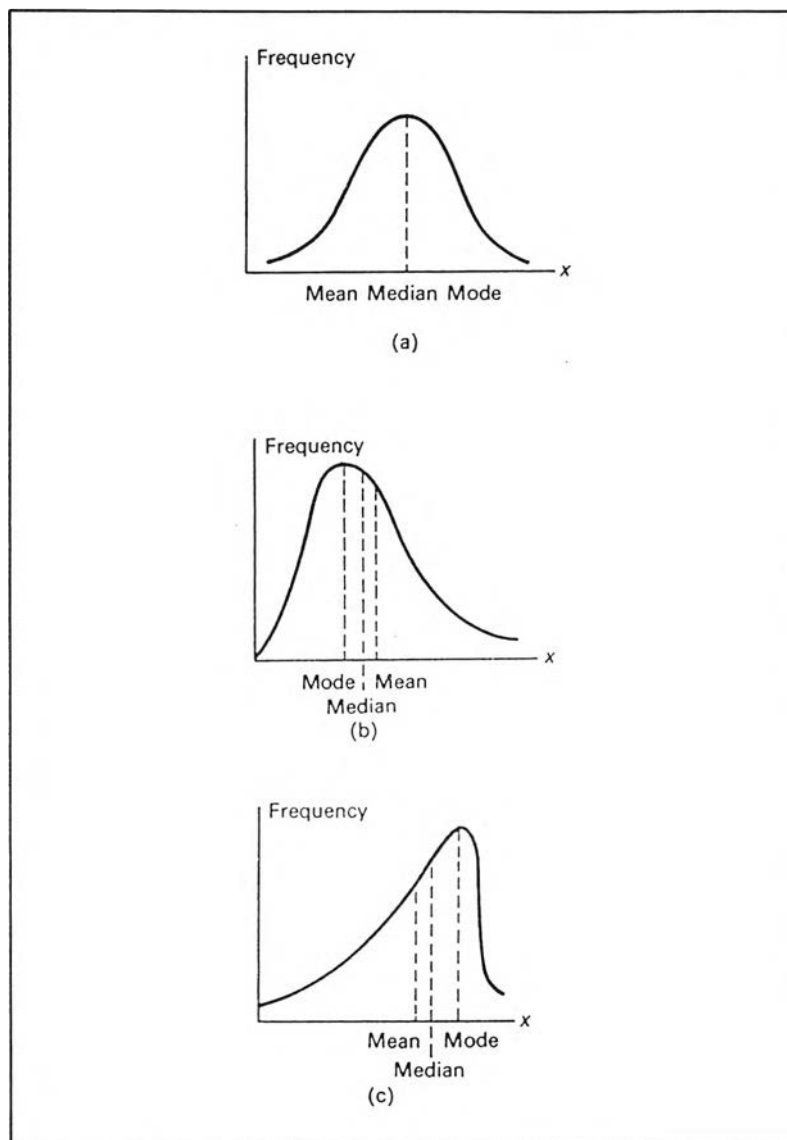
1. ความเบ้ (Skewness) และความโค้ง (Kurtosis)
2. การแจกแจงแลมดาของคูร์
3. การแจกแจงของฟังก์ชันที่ได้จากตัวอย่างสุ่ม และการแจกแจงของค่าเฉลี่ยตัวอย่าง
4. ตัวสถิติที่ใช้ในการทดสอบ

2.1 ความเบ้ (Skewness) และความโค้ง (Kurtosis)

2.1.1 ความเบ้

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

แต่ถ้าประชากรมีการแจกแจงไม่สมมาตร มีลักษณะเบ้ไปข้างใดข้างหนึ่ง ค่าเฉลี่ย มัชยฐาน และฐานนิยมจะมีค่าต่างกัน พิจารณาสรุปต่อไปนี้



รูปที่ 2.1 แสดงเส้นโค้งของการแจกแจงที่ไม่มีควมเบ้ เบ้ขวา และเบ้ซ้าย

จะเห็นว่าในรูป a. ประชากรมีการแจกแจงสมมาตร ซึ่งได้ว่าค่าเฉลี่ย มัธยฐานและฐานนิยม จะมีค่าเท่ากัน ส่วนในรูป b. ประชากรมีการแจกแจงเบ้ไปทางขวา เพราะพื้นที่เส้นโค้งทางด้านขวาของค่าฐานนิยมมากกว่าพื้นที่ทางด้านซ้ายของค่าฐานนิยมและในรูป c. ประชากรมีการแจกแจงเบ้ไปทางซ้าย เพราะพื้นที่ภายใต้เส้นโค้งทางด้านซ้ายของค่าฐานนิยมมีมากกว่าพื้นที่ทางด้านขวาของค่าฐานนิยม

การวัดควมเบ้ (Measure of Skewness) จะใช้การวัดควมเบ้โดยวิธีโมเมนต์ (Moment) สูตรสำหรับหาค่าวัดสมมาตร หรือสัมประสิทธิ์ควมเบ้จากค่าประชากร เป็นดังนี้

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

โดยที่

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

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

สามารถประมาณค่าวัดสมมาตร หรือสัมประสิทธิ์ความเบ้ จากข้อมูลตัวอย่าง ได้ค่าสถิติดังนี้

นี้

$$\alpha_3^\wedge = \frac{m_3}{(m_2)^{3/2}}$$

โดยที่

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

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

การวัดความเบ้ด้วยโมเมนต์ศูนย์กลางอันดับที่ 3 จะให้ค่าต่างๆ กันดังนี้

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

2.1.2 ความโค้ง

การวัดความโค้ง ก็คือการวัดเส้นโค้งว่าจะมีความโค้งมากน้อยเพียงใด เส้นโค้งที่เราเรียกว่าเส้นโค้งปกติ เส้นโค้งใดที่โค้งผิดจากเส้นโค้งปกติ ก็นับเป็นเส้นโค้งที่ไม่ปกติทั้งสิ้น แม้แต่จะเป็นรูปประฆังที่สมมาตรก็ตาม

ความโค้งของการแจกแจงของประชากร มี 3 ลักษณะดังนี้

1. เส้นโค้งที่มีความโค้งเป็นปกติ เรียกว่า เส้นโค้งชนิด Mesokurtic
2. เส้นโค้งที่แบนราบกว่าปกติ เรียกว่า เส้นโค้งชนิด Platykurtic
3. เส้นโค้งที่โค้งกว่าปกติ เรียกว่า เส้นโค้งชนิด Leptokurtic

การวัดความโค้ง (Measure of Kurtosis) หรือการหาสัมประสิทธิ์ความโค้ง 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 $\sqrt{V(X)}$

สามารถประมาณค่าวัดความโค้ง หรือสัมประสิทธิ์ความโค้งจากข้อมูลตัวอย่างได้ค่าสถิติ ดังนี้

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

โดยที่

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

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

การวัดความโค้งโดยใช้โมเมนต์ศูนย์กลางอันดับที่ 4 จะให้ค่าต่างๆ กันดังนี้

1. ถ้า α_4 เท่ากับ 3 แสดงว่าเส้นโค้งมีความโค้งเป็นปกติ (Mesokurtio)
1. ถ้า α_4 น้อยกว่า 3 แสดงว่าเส้นโค้งแบนราบกว่าปกติ (Platykurtio)
3. ถ้า α_4 มากกว่า 3 แสดงว่าเส้นโค้งที่โค้งกว่าปกติ (Leptokurtio)

2.2 การแจกแจงแลมดาของตุกีร์ (Tukey's Lamda Distribution)

Ramberg และ Sohmeiser ได้เสนอวิธีการสร้างตัวแปรสุ่มที่ขึ้นอยู่กับความเบ้ (skewness : α_3) และความโค้ง (kurtosis : α_4) โดยตัวแปรสุ่มนี้มีการแจกแจงที่เรียกว่า “การแจกแจงแลมดาของตุกีร์” โดยที่ตัวแปรสุ่มนั้นจะถูกกำหนดจากค่าพารามิเตอร์ 4 ค่า ซึ่งสัมพันธ์กับค่าความเบ้และความโค้ง ดังนี้

$$X = R(p) = \lambda_1 + [p^{\lambda_3} - (1-p)^{\lambda_3}] / \lambda_2 \quad ; 0 < p < 1 \quad (2.1)$$

โดยที่ p เป็นเลขคู่ที่มีค่าระหว่าง 0 และ 1

λ_1 คือพารามิเตอร์ที่กำหนดตำแหน่ง (Location parameter)

λ_2 คือพารามิเตอร์มาตราส่วน (Scale parameter)

λ_3, λ_4 คือพารามิเตอร์ลักษณะ (Shape parameter) ซึ่งขึ้นกับค่าความเบ้และความโค้งที่กำหนด ถ้าการแจกแจงเป็นแบบสมมาตร จะได้ว่า $\lambda_3 = \lambda_4$

ดังนั้นฟังก์ชันความหนาแน่นของตัวแปรสุ่ม X ที่ได้จากสมการ (2.1) คือ

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

โดยที่

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

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

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

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 ($\mu_3 = E(X - \mu)^3$) และ โมเมนต์ที่ 4 รอบค่าเฉลี่ย ($\mu_4 = E(X - \mu)^4$) จากการแจกแจงได้ดังนี้

$$\mu = \lambda_1 + A / \lambda_2$$

$$\sigma^2 = (B - A^2) / \lambda_2^2$$

$$\mu_3 = (C - 3AB + 2A^3) / \lambda_2^3$$

$$\mu_4 = (D - 4AC + 6A^2B - 3A^4) / \lambda_2^4$$

โดยที่

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

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

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

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

ดังนั้นค่าความเบ้และความโค้ง เป็นไปตามสมการดังนี้

$$\alpha_3 = \mu_3 / \sigma^3 \quad (2.4)$$

และ

$$\alpha_4 = \mu_4 / \sigma^4 \quad (2.5)$$

เราสามารถหาค่า $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ เมื่อกำหนดค่าความเบ้และความโค้งต่างๆ ได้ จากตาราง Remberg แสดงในภาคผนวก โดยที่ค่า เป็นค่าที่ค่าเฉลี่ยเป็นศูนย์และความแปรปรวนเท่ากับ 1 แต่ถ้าเฉลี่ยเท่ากับ μ และค่าความแปรปรวนเท่ากับ σ^2 จะต้องแปลงค่า λ_1 และ λ_2 จากตาราง ดังนี้

$$\lambda_1(\mu, \sigma) = \lambda_1(0,1)\sigma + \mu \quad (2.6)$$

$$\lambda_2(\mu, \sigma) = \lambda_2(0,1) / \sigma \quad (2.7)$$

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

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

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

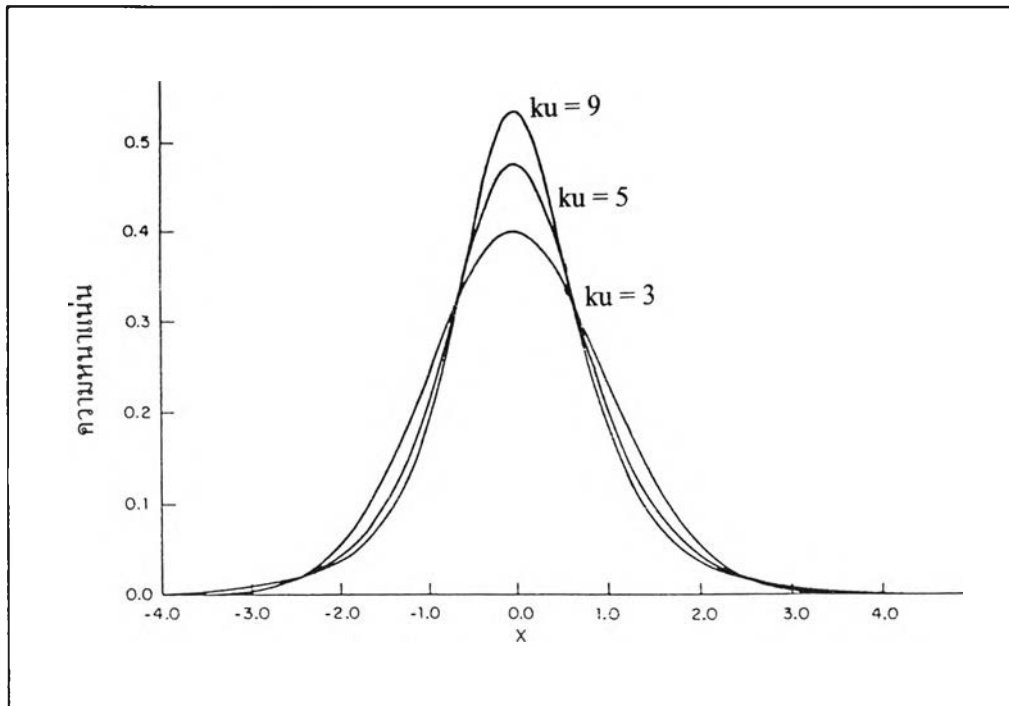
โดยที่

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

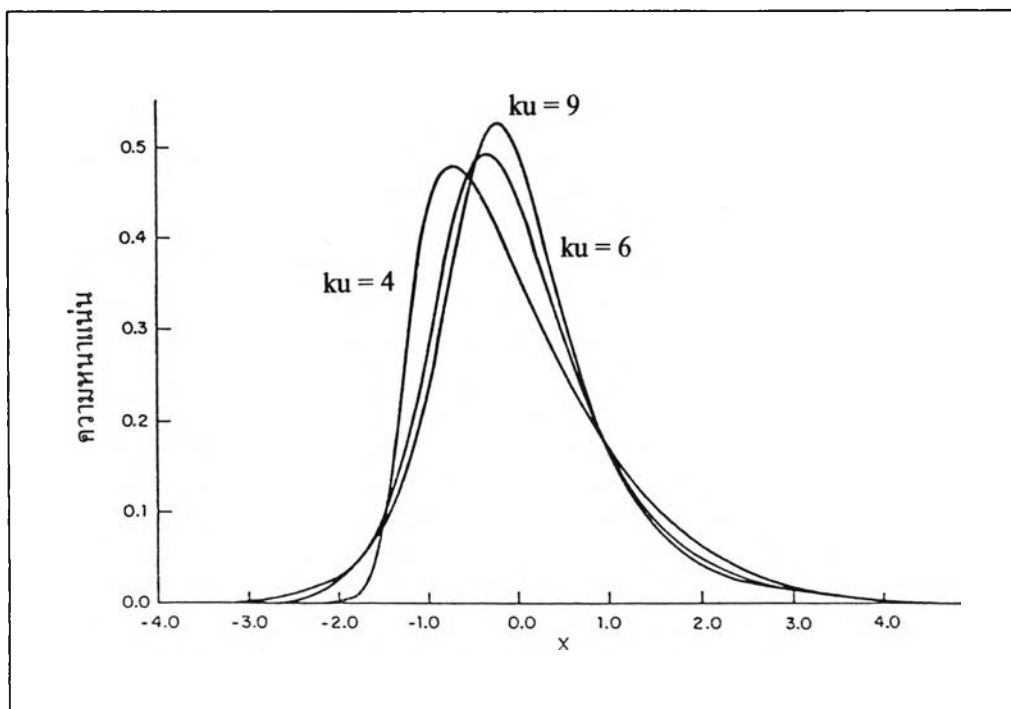
$$m_3 = \sum_{i=1}^n (x_i - \bar{x})^3 / n$$

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

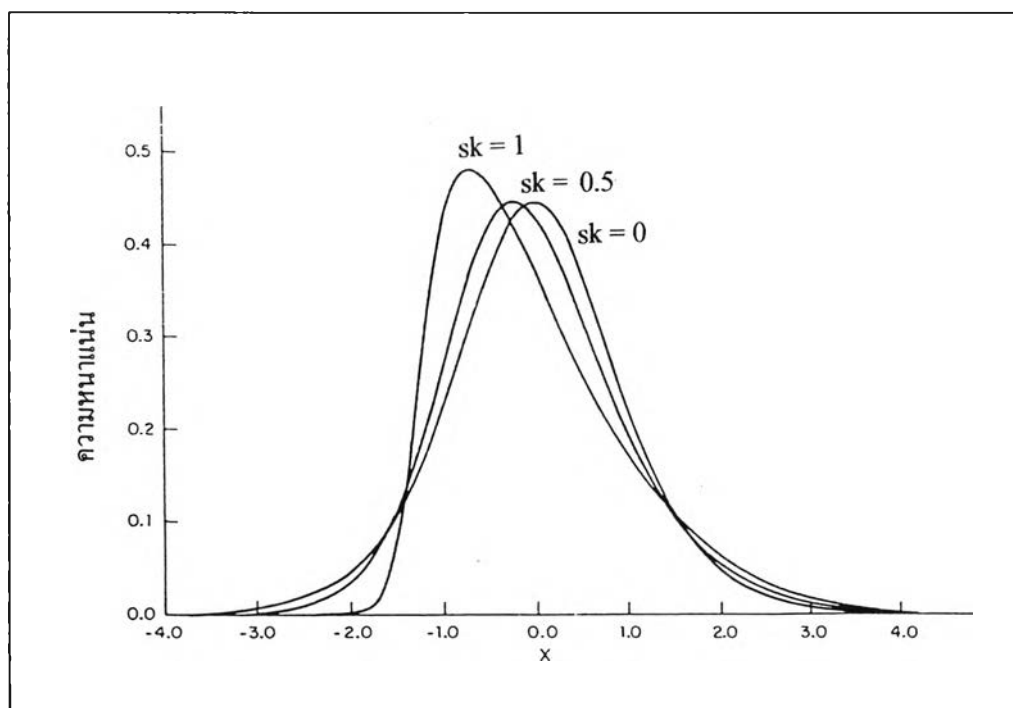
กราฟแสดงฟังก์ชันความหนาแน่นของการแจกแจงแลมดาของคูเกิร์ ที่ค่าความเบ้ และความโค้งต่างๆ แสดงดังรูปที่ 2.2 – 2.4



รูปที่ 2.2 แสดงฟังก์ชันความหนาแน่นของการแจกแจงแลมดาของคูเกิร์ ที่ความเบ้เท่ากับ 0, ความโค้งเท่ากับ 3, 5, 9



รูปที่ 2.3 แสดงฟังก์ชันความหนาแน่นของการแจกแจงแลมดาของคูเกียร์ ที่ความเบ้เท่ากับ 1, ความโค้งเท่ากับ 4, 6, 9



รูปที่ 2.4 แสดงฟังก์ชันความหนาแน่นของการแจกแจงแลมดาของคูเกียร์ ที่ความเบ้เท่ากับ 0, 0.5, 1, ความโค้งเท่ากับ 4

2.3 การแจกแจงของฟังก์ชันที่ได้จากตัวอย่างสุ่ม และการแจกแจงของค่าเฉลี่ยตัวอย่าง

2.3.1 ค่าคาดหวังและค่าความคลาดเคลื่อนมาตรฐานของฟังก์ชันที่ได้จากตัวอย่างสุ่ม

เนื่องจากฟังก์ชันที่ได้จากตัวอย่างสุ่มเป็นตัวแปรสุ่มที่มีการแจกแจงความน่าจะเป็น จึงสามารถที่จะหาค่าคาดหวัง และค่าความคลาดเคลื่อนมาตรฐาน (Standard Error) ของฟังก์ชันหรือตัวแปรสุ่มนี้ได้ สมมติว่า จากตัวอย่างสุ่ม (Y_1, Y_2, \dots, Y_n) ฟังก์ชัน $W = \bar{Y}$ 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, \dots, Y_n) ซึ่งเลือกมาจากประชากรที่มีค่าเฉลี่ย μ และความแปรปรวน σ^2 จะได้ว่าค่าเฉลี่ยตัวอย่าง \bar{Y} จะมีค่าคาดหวังเช่นเดียวกับค่าคาดหวังของประชากร Y และจะมีค่าความแปรปรวนเท่ากับค่าความแปรปรวนของ Y หาคำอธิบายขนาดตัวอย่าง กล่าวคือ

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

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

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

2.3.2 การแจกแจงหาค่าเฉลี่ยตัวอย่าง

2.3.2.1 กรณีไม่ทราบค่าเบี่ยงเบนมาตรฐานของประชากร และประชากรมีการแจกแจงปกติ

ถ้าเลือกตัวอย่างจากประชากรที่มีค่าเฉลี่ย μ และความแปรปรวน σ^2 ค่าเฉลี่ยตัวอย่าง \bar{Y} จะมีการแจกแจงเป็นปกติ โดยมีค่าเฉลี่ย μ และความแปรปรวน σ^2 / n ทำให้ $\frac{\bar{Y} - \mu}{\sigma / \sqrt{n}}$ มีการแจกแจงแบบ $N(0,1)$ อย่างไรก็ตามในทางปฏิบัติบ่อยครั้งเราจะไม่ทราบค่าเบี่ยงเบนมาตรฐานของประชากร ในกรณีเช่นนี้ มักจะประมาณ σ ด้วยค่าเบี่ยงเบนมาตรฐานของตัวอย่าง S ตัวสถิติที่น่าสนใจจึงเป็น $t = \frac{\bar{Y} - \mu}{s / \sqrt{n}}$ ซึ่งไม่ได้มีการแจกแจงปกติอีกต่อไป แต่เป็นตัวสถิติเป็นฟังก์ชันของตัวแปรสุ่ม 2 ตัว คือ \bar{Y} และ S

ดังนั้นตัวสถิติทดสอบ

$$t = \frac{\bar{Y} - \mu}{s / \sqrt{n}}$$

จึงมีการแจกแจงที่ ด้วยระดับชั้นความเสรีเท่ากับ $n-1$

2.3.2.2 กรณีประชากรมีการแจกแจงแบบอื่นๆ ที่ไม่ใช้การแจกแจงปกติ

ในบางครั้งประชากรที่เราสนใจอาจมีการแจกแจงแบบอื่นๆ ที่ไม่ใช้การแจกแจงปกติ โดยปกติเราจะอาศัยทฤษฎีบทลิมิตเข้าสู่ส่วนกลาง เพื่อประมาณการแจกแจงของค่าเฉลี่ยตัวอย่างดังนี้

ถ้าสุ่มตัวอย่างขนาด n (Y_1, Y_2, \dots, Y_n) จากประชากรใดๆ ที่มีค่าเฉลี่ย μ และความแปรปรวน σ^2 แล้ว ถ้าตัวอย่างมีขนาดใหญ่ ค่าเฉลี่ยตัวอย่าง \bar{Y} จะมีการแจกแจงเข้าสู่การแจกแจงปกติที่มีค่าเฉลี่ย μ และความแปรปรวน σ^2/n หรือ $\frac{\bar{Y} - \mu}{\sigma / \sqrt{n}}$ จะมีการแจกแจงโดยประมาณปกติมาตรฐาน แต่ถ้าไม่ทราบค่าเบี่ยงเบนมาตรฐานของประชากร (σ) จะประมาณค่าเบี่ยงเบนมาตรฐานของประชากรด้วยค่าเบี่ยงเบนมาตรฐานของตัวอย่าง (S) ดังนั้นการแจกแจงของ $t = \frac{\bar{Y} - \mu}{s / \sqrt{n}}$ ซึ่งเรียกว่า “ตัวสถิติทดสอบที” จะมีการแจกแจงที่ ถ้าขนาดตัวอย่างมีขนาดใหญ่พอ

แต่ในบางครั้งเราก็ไม่สามารถที่จะสุ่มตัวอย่างให้มีขนาดใหญ่หลายๆ ได้ ดังนั้นตัวสถิติทดสอบที่อาจจะไม่เหมาะสมกับการนำมาใช้งาน เมื่อประชากรมีการแจกแจงแบบอื่น โดยเฉพาะอย่างยิ่งประชากรมีความเบ้ จึงจำเป็นต้องเลือกใช้ตัวสถิติตัวอื่น

2.4 ตัวสถิติที่ใช้ในการทดสอบ

2.4.1 ตัวสถิติทดสอบที (Student's t test)

$$t = \frac{(\bar{x} - \mu_0)}{S / \sqrt{n}}$$

เมื่อ \bar{x} แทนค่าเฉลี่ยของตัวอย่างมีค่าเท่ากับ $\frac{\sum_{i=1}^n x_i}{n}$

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

n แทนขนาดของกลุ่มตัวอย่าง

ตัวสถิติทดสอบที่จะมีการแจกแจงที่ ด้วยระดับชั้นความเสรี (degree of freedom) เป็น n-1 เมื่อสมมติฐานว่างเป็นจริง

จะปฏิเสธสมมติฐานเมื่อ

ทดสอบทางด้านน้อยกว่า : $t < -t_{\alpha, n-1}$

ทดสอบทางด้านมากกว่า : $t > t_{\alpha, n-1}$

ทดสอบสองทาง : $|t| > t_{\frac{\alpha}{2}, n-1}$

2.4.2 ตัวสถิติทดสอบทีของจอห์นสัน (Johnson's t test)

$$J = \left[(\bar{x} - \mu_0) + \frac{\hat{\mu}_3}{6s^2n} + \frac{\hat{\mu}_3}{3s^4} (\bar{x} - \mu_0)^2 \right] / (s/\sqrt{n})$$

เมื่อ \bar{x} แทนค่าเฉลี่ยของตัวอย่างมีค่าเท่ากับ $\frac{\sum_{i=1}^n x_i}{n}$

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

n แทนขนาดของกลุ่มตัวอย่าง

$\hat{\mu}_3$ แทนโมเมนต์ที่สามรอบค่าเฉลี่ยของตัวอย่าง โดย $\hat{\mu}_3 = \sum_{i=1}^n \frac{(x_i - \bar{x})^3}{n}$

จะปฏิเสธสมมติฐานเมื่อ

ทดสอบทางด้านน้อยกว่า : $J < -t_{\alpha, n-1}$

ทดสอบทางด้านมากกว่า : $J > t_{\alpha, n-1}$

ทดสอบสองทาง : $|J| > t_{\frac{\alpha}{2}, n-1}$

2.4.3 ตัวสถิติทดสอบดัดแปลงของจอห์นสัน (Modified Johnson's t test)

$$MJ = \frac{\sqrt{n}(\bar{x} - \mu_0)}{s} + \frac{1}{6\sqrt{n}} \hat{\beta}_1 \left[1 + 2 \left\{ \frac{\sqrt{n}(\bar{x} - \mu_0)}{s} \right\}^2 \right] + \frac{\hat{\beta}_1^2}{9n} \left[\frac{\sqrt{n}(\bar{x} - \mu_0)}{s} + 2 \left\{ \frac{\sqrt{n}(\bar{x} - \mu_0)}{s} \right\}^2 \right]$$

เมื่อ \bar{x} แทนค่าเฉลี่ยของตัวอย่างมีค่าเท่ากับ $\frac{\sum_{i=1}^n x_i}{n}$

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

n แทนขนาดของกลุ่มตัวอย่าง

$\hat{\beta}_1$ แทนตัวประมาณที่ไม่เอนเอียงของค่าความเบี่ยงของตัวอย่างมีค่าเท่ากับ $\frac{n \sum_{i=1}^n (x_i - \bar{x})^3}{(n-1)(n-2)s^3}$

จะปฏิเสธสมมติฐานเมื่อ

ทดสอบทางด้านน้อยกว่า : $MJ < (-t_{\alpha, n-1} - z_\alpha) / 2$

ทดสอบทางด้านมากกว่า : $MJ > (t_{\alpha, n-1} + z_\alpha) / 2$

ทดสอบสองทาง : $|MJ| > \left(t_{\frac{\alpha}{2}, n-1} + \frac{z_\alpha}{2} \right) / 2$