บทที่ 2

ตัวสถิติและผลงานวิจัยที่เกี่ยวข้อง

ตัวสถิติที่ใช้ในการทดสอบ

1. ตัวสถิติ W หรือตัวสถิติของ Shapiro-Wilk

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

เมื่อ n แทน ขนาดตัวอย่าง

k แทน จำนวนเต็มที่เล็กที่สุดที่มากกว่าหรือเท่ากับ $n/2$

a_i แทน ค่าสัมประสิทธิ์ที่ได้จากการเปิดตารางที่ 1 ในภาคผนวก เมื่อ $n \leqslant 50$

$x_{(i)}$ แทน order sample

จะปฏิเสธสมมติฐานว่างที่ว่า "ประชากรมีการแจกแจงแบบปกติ" เมื่อค่าสถิติที่คำนวณได้มีค่าน้อยกว่าค่า W ที่ได้จากตารางในภาคผนวก ที่ขนาดตัวอย่าง n และระดับนัยสำคัญที่ต้องการ

ตัวสถิติ W เป็นตัวสถิติที่ใช้สัดส่วนของ $\hat{\sigma}^2 / s^2$ เมื่อ $\hat{\sigma}^2$ เป็นค่าประมาณของ σ^2 ขั้นตอนในการหาตัวสถิติ W มีดังนี้

กำหนดให้ $m' = (m_1, m_2, \ldots, m_n)$ เป็นเวคเตอร์ค่าคาดหวังของ $x_1, x_2, \ldots, x_n$

$v = (v_{ij})$ เป็น Covariance Matrix ขนาด nxn

$y_1 \leqslant y_2 \leqslant \ldots \leqslant y_n$ เป็น order sample ขนาด n จาก N(0,1)

ดังนั้น $E(x_i) = m_i$

$\mathrm{Cov}(y_i, y_j) = v_{ij} \qquad (i,j = 1,2,\ldots,n)$

ให้ $x' = (x_{(1)}, x_{(2)}, \ldots, x_{(n)})$ เป็นเวคเตอร์ของ order sample ของตัวอย่างขนาด n ต้องการทดสอบว่าตัวอย่างนี้สุ่มมาจากการแจกแจงแบบปกติ $N(\mu, \sigma^2)$

ถ้า $\{x_i\}$ มีการแจกแจงแบบปกติ

จะได้ $$\frac{x_i - \mu}{\sigma} = y_i$$

$$x_i = \mu + \sigma y_i \qquad (i=1,2,\ldots,n)$$

$$E(x_i) = \mu + \sigma E(y_i)$$

$$E(x) = 1 + \sigma m \qquad (1 = \underset{\sim}{1}_{n\times 1})$$

ดังนั้น $$E(x) = p\theta \qquad \ldots\ldots\ldots\ldots(1)$$

เมื่อ p เป็นเมตริกซ์ $(1, m_i)$ ซึ่งมีขนาด nx2

θ' เป็นเวกเตอร์ (μ, σ)

และ $$Var(x_i) = \sigma^2 Var(y_i)$$

$$Var(x) = \sigma^2 v$$

จาก (1) โดยทฤษฎี Generalized Least-Sqrares

$$\hat{\theta} = (p'v^{-1}p)^{-1} p'v^{-1}x \qquad \ldots\ldots\ldots\ldots(2)$$

$p'v^{-1}p$ สามารถกระจายในรูปของ

$$p'v^{-1}p = \begin{bmatrix} 1'v^{-1}1 & 1'v^{-1}m \\ 1'v^{-1}m & m'v^{-1}m \end{bmatrix}$$

$$(p'v^{-1}p)^{-1} = \frac{1}{\Delta} \begin{bmatrix} m'v^{-1}m & -1'v^{-1}m \\ -1'v^{-1}m & 1'v^{-1}1 \end{bmatrix}$$

เมื่อ $$\Delta = 1'v^{-1}1m'v^{-1}m - 1'v^{-1}m1'v^{-1}m$$

$$= 1'v^{-1}1m'v^{-1}m - (1'v^{-1}m)^2$$

ในทำนองเดียวกันสามารถกระจาย $p'v^{-1}x$ ในรูปของ

$$p'v^{-1}x = \begin{bmatrix} 1'v^{-1}x \\ \\ m'v^{-1}x \end{bmatrix}$$

จาก (2) ดังนั้น

$$\hat{\theta} = \frac{1}{\Delta}\begin{bmatrix} m'v^{-1}m & -1'v^{-1}m \\ \\ -1'v^{-1}m & 1'v^{-1}1 \end{bmatrix}\begin{bmatrix} 1'v^{-1}x \\ \\ m'v^{-1}x \end{bmatrix}$$

$$\begin{bmatrix} \hat{\mu} \\ \\ \hat{\sigma} \end{bmatrix} = \frac{1}{\Delta}\begin{bmatrix} m'v^{-1}m1'v^{-1}x - 1'v^{-1}mm'v^{-1}x \\ \\ -1'v^{-1}m1'v^{-1}x + 1'v^{-1}1m'v^{-1}x \end{bmatrix}$$

$$\hat{\mu} = \frac{m'v^{-1}(m1' - 1m')v^{-1}x}{1'v^{-1}1m'v^{-1}m - (1'v^{-1}m)^2}$$

$$\hat{\sigma} = \frac{1'v^{-1}(1m' - m1')v^{-1}x}{1'v^{-1}1m'v^{-1}m - (1'v^{-1}m)^2} \qquad (\because 1'v^{-1}m = 0)$$

$$= \frac{m'v^{-1}x}{m'v^{-1}m}$$

เนื่องจาก $s^2 = \sum_{i=1}^{n} (y_i - \bar{y})^2$ เป็นตัวประมาณที่ไม่เอนเอียงของ $(n-1)\sigma^2$

ดังนั้นตัวสถิติ W ที่ใช้ทดสอบการแจกแจงแบบปกติ คือ

$$W = \frac{R^4 \hat{\sigma}^2}{C^2 S^2} = \frac{b^2}{S^2} = \frac{(a'x)^2}{S^2} = \frac{(\Sigma a_i x_i)^2}{\Sigma (x_i - \bar{x})^2}$$

เมื่อ $R^2 = m'V^{-1}m$

$C^2 = m'V^{-1}V^{-1}m$

$$a' = (a_1, a_2, \ldots, a_m) = \frac{m'V^{-1}}{(m'V^{-1}V^{-1}m)^{1/2}} = \frac{m'V^{-1}}{C}$$

$$b = \frac{R^2 \hat{\sigma}}{C}$$

แต่เนื่องจาก $-a_i = a_{n+1-i}$ จาก Shapiro-Wilk (1965: 593)

ดังนั้น

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

Sarhan และ Greenberg ได้คำนวณค่าของ a' เมื่อขนาดตัวอย่างเล็กกว่า 20 แต่สำหรับขนาดตัวอย่างที่ใหญ่กว่า 20 Shapiro และ Wilk ได้ประมาณค่า a' โดย

$$a' = \frac{m'V^{-1}}{(m'V^{-1}V^{-1}m)^{1/2}}$$

$\therefore \quad a'a = 1$

ให้ $\quad a^* = m'v^{-1}$

ดังนั้น $\quad c^2 = a^*a^{*\prime}$

และประมาณค่า a^* โดย $a_i^* = 2\,m_i$ เมื่อ $i=2,3,\ldots,n-1$

และ
$$(a_1^*)^2 = (a_n^*)^2 = \begin{cases} \dfrac{\tau\{(1/2)n\}}{\sqrt{2\,\tau\{(1/2)(n+1)\}}} & (n \leqslant 20) \\[2ex] \dfrac{\tau\{(1/2)(n+1)\}}{\sqrt{2\,\tau\{(1/2)n+1\}}} & (n > 20) \end{cases}$$

<u>ตัวอย่าง</u> การใช้สถิติของ Shapiro และ Wilk ในการทดสอบการแจกแจงแบบปกติของข้อมูลต่อไปนี้

y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}
-3	-5	4	6.5	0	2	1	3.5	7	10

H_0 : ประชากรมีการแจกแจงแบบปกติ

H_1 : ประชากรไม่มีการแจกแจงแบบปกติ

เรียงลำดับข้อมูลใหม่

i	$x_{(i)}$	$(x_{(i)} - \bar{x})^2$
1	-5	57.76
2	-3	31.36
3	0	6.76
4	1	2.56
5	2	0.36
6	3.5	0.81
7	4	1.96
8	6.5	15.21
9	7	19.36
10	10	54.76
		190.90

$$s^2 = \sum_{i=1}^{10} (x_{(i)} - \bar{x})^2$$

$$= 190.90$$

จากตารางที่ 1 ในภาคผนวก เมื่อ $n=10$

$a_{10} = 0.5739$, $a_9 = 0.3291$, $a_8 = 0.2141$, $a_7 = 0.1224$ และ $a_6 = 0.0399$

$$b = \sum_{i=1}^{k} a_{n+i-1} (x_{(n+i-1)} - x_{(i)})$$

$= (0.5739)(10-(-5)) + (0.3291)(7-(-3)) + (0.2141)(6.5-0)$

$+ (0.1224)(4-1) + (0.0399)(3.5-2)$

$= 13.7182$

$W = b^2 / s^2 = (13.7182)^2 / 190.90 = 0.9858$

จากตารางที่ 2 พบว่า ถ้ากำหนดระดับนัยสำคัญ 0.05 ค่า W ที่ขนาดตัวอย่าง 10 มีค่าเท่ากับ 0.842 ดังนั้นยอมรับ H_o เนื่องจากค่า W ที่คำนวณได้มีค่ามากกว่า 0.842

2. <u>ตัวสถิติ W^2 หรือตัวสถิติของ Cramer-von Mises</u>

$$W^2 = \sum_{i=1}^{n} \{z_i - (2i-1)/2n\}^2 + 1/(12n)$$

เมื่อ n แทน ขนาดตัวอย่าง

z_i แทน ความน่าจะเป็นสะสมของการแจกแจงแบบปกติมาตรฐาน

ซึ่ง $z_i = \phi((x_{(i)} - \bar{x})/s)$

$x_{(i)}$ แทน order sample

$$\bar{x} = \sum_{i=1}^{n} x_{(i)} / n$$

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

สำหรับค่าวิกฤตของตัวสถิติ W^2 จะใช้ค่าจากตารางของ M.A.Stephens (1974) (ตารางที่ 3 ในภาคผนวก) ซึ่งจะต้องมีการปรับค่า W^2 ก่อน เรียกว่า Modified form เขียนแทนด้วย $T^*(W^2) = W^2(1+0.5/n)$ และจะปฏิเสธสมมติฐานว่างเมื่อค่า $T^*(W^2)$ ที่คำนวณได้มากกว่าค่าของ percentage points สำหรับ modified T^* ณ.ระดับนัยสำคัญที่กำหนด

ตัวสถิติ W^2 เป็นตัวสถิติของ Cramer และ Von Mises(1961) ซึ่งได้จากการหาขนาดความแตกต่างระหว่างฟังก์ชันการแจกแจงของตัวอย่าง($F_n(x)$) กับฟังก์ชันการแจกแจงของประชากร($F(x)$) กล่าวคือ

$$J_n = n \int_{-\alpha}^{\alpha} \psi(F(x))(F_n(x) - F(x))^2 \, dF(x)$$

$$= n \int_{0}^{1} \psi(t)(F_n(t) - t)^2 \, dt \qquad \text{เมื่อ } t = F(x)$$

โดย $\psi(t)$ แทน Weight Function

ซึ่งตัวสถิติ W^2 นี้สามารถคำนวณได้จากสมการข้างต้นเมื่อกำหนดให้ $\psi(t) = 1$

<u>ตัวอย่าง</u> การใช้ตัวสถิติของ Cramer-von Mises ในการทดสอบการแจกแจงแบบปกติของข้อมูลต่อไปนี้

y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}
-3	-5	4	6.5	0	2	1	3.5	7	10

H_0 : ประชากรมีการแจกแจงแบบปกติ

H_1 : ประชากรไม่มีการแจกแจงแบบปกติ

เรียงลำดับข้อมูลใหม่

i	$x_{(i)}$	$(x_{(i)}-\bar{x})/s$	z_i	$(2i-1)/2n$	$(z_i-(2i-1)/2n)^2$
1	-5	-1.65	0.0495	0.05	2.50×10^{-7}
2	-3	-1.22	0.1112	0.15	1.51×10^{-3}
3	0	-0.56	0.2877	0.25	1.42×10^{-3}
4	1	-0.34	0.3632	0.35	1.74×10^{-4}
5	2	-0.13	0.4483	0.45	2.89×10^{-6}
6	3.5	0.19	0.5793	0.55	8.58×10^{-4}
7	4	0.30	0.6179	0.65	1.03×10^{-3}
8	6.5	0.85	0.8023	0.75	2.74×10^{-3}
9	7	0.96	0.8315	0.85	3.42×10^{-4}
10	10	1.61	0.9463	0.95	1.37×10^{-5}
					8.08×10^{-3}

$$\bar{x} = \sum_{i=1}^{10} x_{(i)} / 10 = 2.60$$

$$s^2 = \sum_{i=1}^{10} (x_{(i)} - \bar{x})^2 / 10-1 = 4.6005$$

$$W^2 = \sum_{i=1}^{n} \{z_i - (2i-1)/2n\}^2 + 1/(12n)$$

$$= (8.08\times10^{-3}) + (1/(12)(10))$$

$$= 0.0164$$

ปรับค่า W^2 ให้อยู่ในรูป Modified form ได้ดังนี้

$$T^*(W^2) = W^2(1+0.5/n)$$

$$= (0.0164)(1+0.5/10)$$

$$= 0.0172$$

จากตารางที่ 3 ในภาคผนวกพบว่า ถ้ากำหนดระดับนัยสำคัญ 0.05 ค่า $T^*(W^2)$ มีค่าเท่ากับ 0.126 ดังนั้นยอมรับ H_0 นั่นคือ ประชากรมีการแจกแจงแบบปกติ

3. <u>ตัวสถิติ A^2 หรือตัวสถิติของ Anderson-Darling</u>

$$A^2 = -n - [\sum_{i=1}^{n} (2i-1) \{\ln z_i + \ln (1-z_{n+1-i})\} / n]$$

เมื่อ n แทน ขนาดตัวอย่าง

z_i แทน ความน่าจะเป็นสะสมของการแจกแจงแบบปกติมาตรฐาน

ซึ่ง $z_i = \phi((x_{(i)} - \bar{x})/s)$

$x_{(i)}$ แทน order sample

$$\bar{x} = \sum_{i=1}^{n} x_{(i)} / n$$

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

ในทำนองเดียวกันค่าวิกฤตของตัวสถิติ A^2 ใช้ค่าที่ปรับปรุงแล้ว โดยหาจากตารางของ M.A.Stephens เช่นเดียวกัน สำหรับค่าที่ปรับปรุงแล้วของตัวสถิติ A^2 เขียนแทนด้วย $T^*(A^2)$ โดย $T^*(A^2) = A^2(1+4/n-25/n^2)$ และจะปฏิเสธสมมติฐานว่างเมื่อค่า $T^*(A^2)$ ที่คำนวณได้มีค่ามากกว่าค่าของ percentage points สำหรับ modified T^* ณ. ระดับนัยสำคัญที่กำหนด

ตัวสถิติ A^2 เป็นตัวสถิติที่หามาโดย Anderson-Darling (1963) ซึ่งคำนวณได้จากการกำหนด weight function สำหรับสมการ

$$J_n = n \int_0^1 \psi(t)(F_n(t)-t)^2 dt$$

เป็น $\psi(t) = 1/t(t-1)$

<u>ตัวอย่าง</u> การใช้ตัวสถิติของ Anderson-Darling ในการทดสอบการแจกแจงแบบปกติของข้อมูลต่อไปนี้

y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}
-3	-5	4	6.5	0	2	1	3.5	7	10

H_0 : ประชากรมีการแจกแจงแบบปกติ

H_1 : ประชากรไม่มีการแจกแจงแบบปกติ

เรียงลำดับข้อมูลใหม่

017542

i	$x_{(i)}$	z_i	$\ln z_i$	$\ln(1-z_{n+i-i})$	$\ln z_i+\ln(1-z_{n+i-i})$	$2i-1$
1	-5	0.0495	-3.0262	-2.9243	-5.9505	1
2	-3	0.1112	-2.1893	-1.7808	-3.9700	3
3	0	0.2877	-1.2458	-1.6210	-2.8668	5
4	1	0.3632	-1.0128	-0.9621	-1.9748	7
5	2	0.4483	-0.8023	-0.8658	-1.6681	9
6	3.5	0.5793	-0.5459	-0.5947	-1.1406	11
7	4	0.6179	-0.4814	-0.4513	-0.9327	13
8	6.5	0.8023	-0.2203	-0.3393	-0.5595	15
9	7	0.8315	-0.1845	-0.1178	-0.3024	17
10	10	0.9463	-0.0552	-0.0507	-0.1059	19

i	$(2i-1)(\ln z_i+\ln(1-z_{n+1-i}))$
1	-5.9505
2	-11.9100
3	-14.3340
4	-13.8236
5	-15.0129
6	-12.5466
7	-12.1251
8	-8.3925
9	-5.1428
10	-2.0121
	-101.2481

$$A^2 = -n - [\sum_{i=1}^{n} (2i-1)\{\ln z_i + \ln(1-z_{n+1-i})\} / n]$$

$$= -10-(-101.2481)/10$$

$$= 0.1281$$

ทำการปรับให้อยู่ในรูป Modified form ได้ดังนี้

$$T^*(A^2) = A^2(1+4/n-25/n^2)$$

$$= (0.1281)(1+4/10-25/100)$$

$$= 0.1473$$

ค่า $T^*(A^2)$ จากตารางที่ 3 ในภาคผนวกเท่ากับ 0.787 ที่ระดับนัยสำคัญ 0.05 ดังนั้น

ยอมรับ H_o นั่นคือ ประชากรมีการแจกแจงแบบปกติ

4. <u>ตัวสถิติ U^2 หรือตัวสถิติของ Watson</u>

$$U^2 = W^2 - n(\bar{z} - 1/2)^2$$

เมื่อ n แทน ขนาดตัวอย่าง

W^2 แทน ตัวสถิติของ Cramer-von Mises

$$\bar{z} = \sum_{i=1}^{n} z_i / n$$

z_i แทน ความน่าจะเป็นสะสมของการแจกแจงแบบปกติมาตรฐาน

ซึ่ง $z_i = \phi((x_{(i)} - \bar{x})/s)$

$x_{(i)}$ แทน order sample

$$\bar{x} = \sum_{i=1}^{n} x_{(i)} / n$$

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

ในทำนองเดียวกันเมื่อปรับตัวสถิติ U^2 ให้อยู่ในรูป Modified form โดย $T^*(U^2) = U^2(1+0.5/n)$ จะปฏิเสธสมมติฐานว่างเมื่อค่าที่คำนวณได้มีค่ามากกว่าค่าของ percentage points สำหรับ Modified form T^* ณ.ระดับนัยสำคัญที่กำหนด โดยใช้ตารางของ M.A.Stephens เช่นเดียวกัน

ตัวสถิติ U^2 เป็นตัวสถิติที่ Watson หามาในปี ค.ศ.1962 โดยปรับปรุงมาจากตัวสถิติของ Cramer-von Mises ซึ่งได้จากการอินทิเกรตฟังก์ชัน

$$U^2 = n\int_{-\alpha}^{\alpha} \{F_n(x) - F(x) - \int_{-\alpha}^{\alpha} (F_n(x) - F(x))\, dF(x)\}^2\, dF(x)$$

แล้วจัดให้อยู่ในรูปแบบที่เหมาะสมต่อการนำไปใช้งาน

<u>ตัวอย่าง</u> การใช้ตัวสถิติของ Watson ในการทดสอบการแจกแจงแบบปกติของข้อมูลต่อไปนี้

y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}
-3	-5	4	6.5	0	2	1	3.5	7	10

H_0 : ประชากรมีการแจกแจงแบบปกติ

H_1 : ประชากรไม่มีการแจกแจงแบบปกติ

เรียงลำดับข้อมูลใหม่

i	$x_{(i)}$	z_i
1	-5	0.0495
2	-3	0.1112
3	0	0.2877
4	1	0.3632
5	2	0.4483
6	3.5	0.5793
7	4	0.6197
8	6.5	0.8023
9	7	0.8315
10	10	0.9463
		5.0372

$$z = \sum_{i=1}^{n} z_i / n = 5.0372/10 = 0.50372$$

$$U^2 = W^2 - n(\bar{z} - 1/2)^2$$ (จากตัวอย่างที่ 2 ค่า $W^2 = 0.0164$)

$$= 0.0164 - 10(5.0372/10 - 1/2)^2$$

$$= 0.0163$$

จัดให้อยู่ในรูป Modified form ได้ดังนี้

$$T^*(U^2) = U^2(1+0.5/n)$$

$$= (0.0163)(1+0.5/10)$$

$$= 0.0171$$

ค่า $T^*(U^2)$ จากตารางที่ 3 ในภาคผนวกมีค่าเท่ากับ 0.116 ที่ระดับนัยสำคัญ 0.05 ดังนั้นยอมรับ H_o นั่นคือ ประชากรมีการแจกแจงแบบปกติ

5. <u>ตัวสถิติ V หรือตัวสถิติของ Kuiper</u>

$$V = \max_{1 \leq i \leq n} (i/n - z_i) + \max_{1 \leq i \leq n} (z_i - (i-1)/n)$$

เมื่อ n แทน ขนาดตัวอย่าง

z_i แทน ความน่าจะเป็นสะสมของการแจกแจงแบบปกติมาตรฐาน

ซึ่ง $z_i = \phi((x_{(i)} - \bar{x})/s)$

$x_{(i)}$ แทน order sample

$$\bar{x} = \sum_{i=1}^{n} x_{(i)} / n$$

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

สำหรับค่าวิกฤตของตัวสถิติ V ต้องมีการปรับตัวสถิติให้อยู่ในรูป Modified form เช่นเดียวกัน ซึ่ง $T^*(V) = V(\sqrt{n} + 0.05 + 0.82/\sqrt{n})$ จะปฏิเสธสมมติฐานเมื่อค่าที่คำนวณได้มีค่ามากกว่า percentage points สำหรับ $T^*(V)$ ณ.ระดับนัยสำคัญที่กำหนด โดยใช้ตารางของ M.A.Stephens เช่นเดียวกัน

<u>ตัวอย่าง</u> การใช้ตัวสถิติของ Kuiper ในการทดสอบการแจกแจงแบบปกติของข้อมูลต่อไปนี้

y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}
-3	-5	4	6.5	0	2	1	3.5	7	10

H_0 : ประชากรมีการแจกแจงแบบปกติ

H_1 : ประชากรไม่มีการแจกแจงแบบปกติ

เรียงลำดับข้อมูลใหม่

i	$x_{(i)}$	z_i	i/n	$i/n-z_i$	(i-1)/n	$z_i-(i-1)/n$
1	-5	0.0495	0.1	0.0505	0.0	0.0495
2	-3	0.1112	0.2	0.0888	0.1	0.0112
3	0	0.2877	0.3	0.0123	0.2	0.0877
4	1	0.3632	0.4	0.0368	0.3	0.0632
5	2	0.4483	0.5	0.0517	0.4	0.0483
6	3.5	0.5793	0.6	0.0207	0.5	0.0793
7	4	0.6179	0.7	0.0821	0.6	0.0179
8	6.5	0.8023	0.8	-0.0023	0.7	0.1023
9	7	0.8315	0.9	0.0685	0.8	0.0315
10	10	0.9463	1.0	0.0537	0.9	0.0463

$$V = \max_{1\leq i\leq n} (i/n - z_i) + \max_{1\leq i\leq n} (z_i - (i-1)/n)$$

$$= 0.0888 + 0.1023$$

$$= 0.1911$$

เปลี่ยนให้อยู่ในรูป Modified form ได้ดังนี้

$$T^*(V) = V(\sqrt{n} + 0.05 + 0.82/\sqrt{n})$$

$$= (0.1911)(\sqrt{10} + 0.05 + 0.8/\sqrt{10})$$

$$= 0.6634$$

จากตารางที่ 3 ในภาคผนวกพบว่า ค่า $T^*(V)$ มีค่าเท่ากับ 1.489 ที่ระดับนัยสำคัญ 0.05 ดังนั้นยอมรับ H_o นั่นคือ ประชากรมีการแจกแจงแบบปกติ

6. <u>ตัวสถิติ D หรือตัวสถิติของ Durbin</u>

$$D = \max_{1\leq i\leq n} (i/n - \sum_{j=1}^{i} g_j)$$

เมื่อ n แทน ขนาดตัวอย่าง

$$g_j = (n+2-j)(c_{(j)} - c_{(j-1)}) \qquad j=1,2,\ldots,n$$

และ $0 < c_{(0)} < c_{(1)} < c_{(2)} < \ldots < c_{(n)}$ ซึ่งได้มาจากการจัดอันดับค่า $c_0, c_1, c_2, \ldots, c_n$

ซึ่ง $c_0 = z_1$

$c_1 = z_2 - z_1$

$\vdots$

$c_n = 1 - z_n$

โดย z_i แทน ความน่าจะเป็นสะสมของการแจกแจงแบบปกติมาตรฐาน

ซึ่ง $z_i = \phi((x_{(i)} - \bar{x})/s)$

$x_{(i)}$ แทน order sample

$$\bar{x} = \sum_{i=1}^{n} x_{(i)} / n$$

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

สำหรับค่าวิกฤตของตัวสถิติ D ใช้ค่าจากตารางที่ 4 ในภาคผนวก จะปฏิเสธสมมติฐานว่าง ถ้าค่าสถิติที่คำนวณได้มีค่ามากกว่าค่า D ที่เปิดจากตารางที่ขนาดตัวอย่าง n และระดับนัยสำคัญที่ต้องการ

ตัวสถิติ D เป็นตัวสถิติที่คิดขึ้นโดย J.Durbin (1962) โดยปรับปรุงมาจากตัวสถิติของ Kolmogorov-Smirnov ดังนั้นในบางครั้งจึงถูกเรียกว่าเป็น Modified Kolmogorov Test ซึ่งการคำนวณตัวสถิติ D นี้ยังคงใช้หลักการพิจารณาขนาดของความแตกต่างระหว่างฟังก์ชันการแจกแจงของตัวอย่างกับฟังก์ชันการแจกแจงของประชากรเช่นเดียวกัน

<u>ตัวอย่าง</u> การใช้ตัวสถิติของ Durbin ในการทดสอบการแจกแจงแบบปกติของข้อมูลต่อไปนี้

y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}
-3	-5	4	6.5	0	2	1	3.5	7	10

H_0 : ประชากรมีการแจกแจงแบบปกติ

H_1 : ประชากรไม่มีการแจกแจงแบบปกติ

เรียงลำดับข้อมูลใหม่

i	$x_{(i)}$	z_i	i/n	$\sum_{j=1}^{i} g_j$	$i/n-\sum_{j=1}^{i} g_j$
1	-5	0.0495	0.1	0.1232	-0.0232
2	-3	0.1112	0.2	0.2142	-0.0142
3	0	0.2877	0.3	0.2520	0.0480
4	1	0.3632	0.4	0.3160	0.0840
5	2	0.4483	0.5	0.4126	0.0874
6	3.5	0.5793	0.6	0.4702	0.1298
7	4	0.6179	0.7	0.6187	0.0813
8	6.5	0.8023	0.8	0.6835	0.1165
9	7	0.8315	0.9	0.8200	0.0800
10	10	0.9463	1.0	0.8358	0.1642

$$c_0 = z_1 = 0.0495$$

$$c_1 = z_2 - z_1 = 0.0617$$

$$c_2 = z_3 - z_2 = 0.1765$$

$$c_3 = z_4 - z_3 = 0.0755$$

$$c_4 = z_5 - z_4 = 0.0851$$

$$c_5 = z_6 - z_5 = 0.1310$$

$$c_6 = z_7 - z_6 = 0.0404$$

$$c_7 = z_8 - z_7 = 0.1844$$

$$c_8 = z_9 - z_8 = 0.0292$$

$$c_9 = z_{10} - z_9 = 0.1148$$

$$c_{10} = 1 - z_{10} = 0.0537$$

จัดอันดับ $c_0, c_1, \ldots, c_{10}$ ได้ดังนี้

$c_{(0)} = 0.0292$, $c_{(1)} = 0.0404$, $c_{(2)} = 0.0495$, $c_{(3)} = 0.0537$

$c_{(4)} = 0.0617$, $c_{(5)} = 0.0755$, $c_{(6)} = 0.0851$, $c_{(7)} = 0.1148$

$c_{(8)} = 0.1310$, $c_{(9)} = 0.1765$, $c_{(10)} = 0.1844$

คำนวณค่า $g_j = (n+2-j)(c_{(j)} - c_{(j-1)})$ $j=1,2,\ldots,n$ ได้ดังนี้

j	$n+2-j$	$c_{(j)} - c_{(j-1)}$	g_j
1	11	0.0112	0.1232
2	10	0.0091	0.0910
3	9	0.0042	0.0378
4	8	0.0080	0.0640
5	7	0.0138	0.0966
6	6	0.0096	0.0576
7	5	0.0297	0.1485
8	4	0.0162	0.0648
9	3	0.0455	0.1365
10	2	0.0079	0.0158

$$D = \max_{1 \leq i \leq n} \left(i/n - \sum_{j=1}^{i} g_j\right)$$

$$= 0.1642$$

จากตารางที่ 4 ในภาคผนวก พบว่าค่า D เท่ากับ 0.208 ที่ระดับนัยสำคัญ 0.05 ดังนั้น ยอมรับ H_0 นั่นคือ ประชากรมีการแจกแจงแบบปกติ

ตัวสถิติอื่นบางตัวสำหรับทดสอบการแจกแจงแบบปกติ

เนื่องจากตัวสถิติสำหรับทดสอบการแจกแจงแบบปกติมีเป็นจำนวนมาก ในส่วนนี้จะเสนอตัวสถิติเพียงบางตัวเท่านั้น โดยจะเสนอเฉพาะตัวสถิติที่นิยมใช้กันมาก 3 ตัว คือ

1. การทดสอบความโด่ง (Test of Kurtosis)

$$g_2 = b_2 - 3 = (m_4 / m_2^2) - 3$$

เมื่อ $m_k = \sum_{i=1}^{n} (x_i - \bar{x})^k / n$

การหาค่าวิกฤตทำได้โดยใช้การแจกแจงแบบปกติประมาณการแจกแจงของ g_2 ซึ่งมีค่าเฉลี่ยเป็น 0 และส่วนเบี่ยงเบนมาตรฐานเป็น $\sqrt{24/n}$ เมื่อขนาดตัวอย่างมากกว่า 1,000 ในกรณีที่ขนาดตัวอย่างมีค่าน้อย กล่าวคือมีค่าตั้งแต่ 50 ถึง 1,000 ค่า g_2 ที่ระดับนัยสำคัญ 0.05 และ 0.01 หาจากตารางที่ 5 ในภาคผนวก

2. การทดสอบความเบ้ (Test of Skewness)

$$\sqrt{b_1} = g_1 = m_3 / (m_2)^{3/2}$$

เมื่อ $m_k = \sum_{i=1}^{n} (x_i - \bar{x})^k / n$

ภายใต้การแจกแจงแบบปกติ ค่า g_1 จะลู่เข้าสู่ศูนย์ ในการตัดสินใจว่า g_1 แตกต่างจากศูนย์อย่างมีนัยสำคัญหรือไม่ ใช้การแจกแจงแบบปกติประมาณการแจกแจงของ g_1 ซึ่งมีค่าเฉลี่ยเป็น 0 และส่วนเบี่ยงเบนมาตรฐานเป็น $\sqrt{6/n}$ การทดสอบนี้จะถูกต้องเมื่อตัวอย่างมีขนาดตั้งแต่ 150 ขึ้นไป สำหรับ n = 25 ถึง 200 ค่า g_1 มีนัยสำคัญที่ระดับ 0.05 และ 0.01 ในการทดสอบทางเดียว ใช้ตารางที่ 6 ในภาคผนวก

3. ตัวสถิติ D'Agostino (D'Agostino Statistic)

D'Agostino (1971) ได้สร้างตัวสถิติ D ขึ้นโดยแปลงจากตัวสถิติของ Shapiro และ Wilk กล่าวคือ

$$D = T / n^2S$$

เมื่อ $T = \sum_{i=1}^{n} \{i-(1/2)(n-1)\} x_{(i)}$

ภายใต้การแจกแจงแบบปกติ

$E(D) = 1/(2\sqrt{\pi})$

$SD.(D) = 0.02998598/\sqrt{n}$

ดังนั้น $y = \{D - (2\sqrt{\pi})^{-1}\} / S.D. \sim N(0,1)$

จากผลการศึกษาของ D'Agostino พบว่า ถ้าการแจกแจงมีความโด่งน้อยกว่า 3 ค่าของ y จะมากกว่า 0 และถ้าการแจกแจงมีความโด่งมากกว่า 3 ค่าของ y จะน้อยกว่า 0

ผลงานวิจัยที่เกี่ยวข้อง

สำหรับงานวิจัยที่เกี่ยวข้องนั้น ได้มีนักสถิติจำนวนมากทำการศึกษาเกี่ยวกับอำนาจการทดสอบของตัวสถิติที่ใช้ในการทดสอบการแจกแจงแบบปกติ ซึ่งในส่วนนี้จะเสนอเฉพาะบางผลงานวิจัยที่สำคัญและเกี่ยวข้องกับวิทยานิพนธ์เรื่องนี้เท่านั้น

Shapiro และคณะ(1968) เป็นผู้ริเริ่มในการศึกษาอำนาจการทดสอบของตัวสถิติต่าง ๆ ในการทดสอบการแจกแจงแบบปกติ โดยทำการศึกษาตัวสถิติ 9 ตัว คือ Shapiro-Wilk Statistic(W) $\sqrt{b_1}$ b_2 Kolmogorov-Smirnov Test(K) Cramer-von Mises(W^2) Anderson-Darling(A^2) Durbin(D) Chi-sqrare Test(X^2) และ Studentized Range Test(U) ภายใต้การแจกแจง 12 การแจกแจง ซึ่งมีค่าพารามิเตอร์ที่แตกต่างกันรวมเป็น 45 การแจกแจง ได้ผลสรุปดังนี้

1. Shapiro-Wilk Statistic ใช้ได้ดีในการทดสอบทั่วไป

2. การทดสอบโดยใช้ Empirical Distribution Function ได้อำนาจการทดสอบต่ำ

3. Studentized Range Test (U) มีอำนาจการทดสอบสูงเมื่อประชากรมีการแจกแจงแบบ Symmetric Short-Tailed และมีอำนาจการทดสอบต่ำเมื่อประชากรมี

การแจกแจงแบบ Asymmetric Short-Tailed และ Asymmetric Long-Tailed

4. $\sqrt{b_1}$ และ b_2 ใช้ในการทดสอบได้ดี แต่มีอำนาจการทดสอบต่ำกว่า W

M.A.Stephens (1974) ได้แสดงให้เห็นว่า ผลการศึกษาของ Shapiro และคณะ ได้ให้ค่าวิกฤตสำหรับตัวสถิติที่ใช้ Empirical Distribution Function ในการทดสอบการแจกแจงแบบปกติของประชากรที่ไม่ถูกต้องนัก ทั้งนี้เพราะค่าวิกฤตดังกล่าวคำนวณมาจากการสมมติว่าทราบค่าเฉลี่ยและความแปรปรวนของประชากร ดังนั้น Stephens จึงทำการคำนวณค่าวิกฤตของตัวสถิติเหล่านี้ขึ้นมาใหม่ โดยสมมติว่าไม่ทราบค่าเฉลี่ยและความแปรปรวนของประชากร ซึ่งแสดงอยู่ในตารางที่ 3 ในภาคผนวก

สมพิศ โชติวิทยะธารากร (2531) ได้ทำการเปรียบเทียบอำนาจการทดสอบของตัวสถิติที่ใช้ในการทดสอบการแจกแจงแบบปกติ 5 ตัว คือ Chi-Sqrare Test(X^2) Studentized Range Test(U) Shapiro-Wilk Statistic(W) Probability Plot Correlation Coefficient Test(r) และ Hannu oja Statistic(T_1 and T_2) ภายใต้การแจกแจง 2 ลักษณะที่สำคัญ คือ การแจกแจงแบบปกติ และการแจกแจงแบบไม่ปกติ ซึ่งพบว่าในด้านการนำไปใช้ประโยชน์แล้วควรเลือกใช้ตัวสถิติ W เนื่องจากให้อำนาจการทดสอบสูงเป็นส่วนใหญ่ และสามารถควบคุมความคลาดเคลื่อนชนิดที่ 1 ได้ดี